Retrieval Practice Can Promote New Learning With Both Related and Unrelated Prose Materials

Verena M. Kriechbaum, Karl-Heinz T. Bäuml

Department of Experimental Psychology, Regensburg University, Germany

To appear in:

Journal of Applied Research in Memory and Cognition

Correspondence Address:

Karl-Heinz T. Bäuml

Department of Experimental Psychology, Regensburg University,

93040 Regensburg, Germany.

Email: karl-heinz.baeuml@ur.de

Phone: +49-941-943-3818

Abstract

Retrieval practice of studied information can facilitate recall of subsequently encountered, new information, relative to both restudy and other no-retrieval-practice conditions. Here, this forward testing effect was examined in three experiments employing both related and unrelated prose passages as study material. Participants studied five prose passages and were tested on the final passage. Between study of the single passages, participants practiced retrieval of the immediately preceding passage, restudied the passage, generated semantic information unrelated to the single passages, or conducted simple arithmetic tasks. The forward testing effect arose both when the passages were related and when they were unrelated. However, only with unrelated passages did the effect generalize to semantic generation, and only with related passages did retrieval practice create recall superior in level to recall when the final passage was studied only. Relatedness of materials seems to influence how retrieval practice promotes new learning.

Keywords

retrieval practice – testing effect – forward testing effect – prose material

General Audience Summary

Students, during exam preparation, sometimes face the study of quite unrelated materials relevant to different courses and at other times the study of highly related materials relevant to a single course. In both situations, students need to master vast amounts of complex material and clues would thus be helpful on how to optimize new learning. Here, we addressed the issue of whether (i) retrieval practice interspersed between the study of such materials can aid learning and memory of the final critical material, and (ii) such beneficial effect of retrieval practice would be influenced by the relatedness of study materials. The results of three experiments are reported, in which participants were presented with five unrelated or five related prose passages on educationally relevant topics, and were tested on the final critical passage immediately after its presentation. In all experiments, recall of the final passage was higher when, between presentation of the single passages, participants engaged in retrieval practice compared to restudy of the preceding passages, thus demonstrating a recall benefit in response to retrieval practice with both types of prose material. Strikingly, only with unrelated prose passages did a similar recall benefit arise when participants engaged in retrieval of other (semantic) information between the single passages, and only with related prose passages was recall of the final passage higher when the preceding passages were presented and retrieved than when no preceding passages were presented. These findings suggest that retrieval practice can promote new learning in both putative courses scenarios entertained above. However, intermittent retrieval of information unrelated to the study material (e.g. setting up the evening grocery list) may promote new learning only when study materials refer to different courses, and skipping preceding materials may impair learning of the final material only when study materials refer to a single course.

Retrieval Practice Can Promote New Learning With Both Related and Unrelated Prose Materials

Courses are a central part of a student`s educational experience. Students of psychology, for instance, often participate in several courses during a semester, like courses on clinical psychology, organizational psychology, or educational psychology. At the end of the semester, when preparing for the final exams, the students may therefore during some weeks have to study materials relevant to different courses, and during other weeks highly related materials relevant to a single course. In both scenarios, the question arises of how the students can optimize new learning considering the vast amount of study material that will be tested. A wealth of research has established that a powerful tool to promote learning of new information is active retrieval practice.

Indeed, studies on the so-called testing effect have shown that retrieval practice on studied material can improve its long-term retention much more than restudy of the same material (e.g., Roediger & Karpicke, 2006), a finding that has turned out to be very robust and applicable in educational practice (Dunlosky et al., 2013; Roediger & Butler, 2011). However, retrieval practice can also promote learning and memory of subsequently encountered, new material. Szpunar et al. (2008) demonstrated the effect in multiple-list learning, using words as study material. Participants studied five lists of words in anticipation of a final cumulative recall test and, immediately after study of lists 1 through 4, were asked to recall the words of the preceding list (retrieval-practice condition), study the words once again (restudy condition), or solve simple arithmetic tasks (distractor condition). At test, subjects in the retrieval-practice condition recalled more words from critical list 5 and showed fewer intrusions of words from lists 1 through 4 than participants in the other two conditions. This beneficial effect of interim retrieval practice, be it relative to restudy or other no-retrieval-practice conditions, is often referred to as the forward testing (FT) effect.

The FT Effect With Complex Study Material

The FT effect has been shown to arise across a variety of study materials (see Pastötter & Bäuml, 2014; Yang et al., 2018), but research on the effect with more complex study materials, such as video lectures or prose passages, is relatively scarce to date (for early demonstrations in similar paradigms that answering questions during reading of prose materials can facilitate learning, see e.g., Frase, 1968; Rothkopf, 1966). Thus, providing clues for students on how to optimize new learning when, during exam preparation, they face the study of rather unrelated materials or the study of highly related materials may be premature.

To the best of our knowledge, only three experiments published in three separate studies addressed the FT effect to date employing unrelated complex materials. One study used two unrelated video lectures ("Star Life Cycle" and "Lightning Formation") and found an FT effect for the second lecture when subjects engaged in retrieval practice compared to restudy of the first lecture (Yue et al., 2015, Experiment 2). In the second study, participants were presented with three expository texts on a common topic ("Forms of Government Intervention in the U.S. Labor Market"), but the three texts ("Benefit Mandates", "Labor Laws", "Job Training Programs") were not directly related to one another and had no overlapping information; results revealed an FT effect relative to a no-retrieval-practice baseline (Wissman et al., 2011, Experiment 1A; see also Wissman & Rawson, 2015). The third study employed completely unrelated prose passages as study materials (e.g., "Porcupines", "Coyotes", "Chronic Wasting Disease") but did not directly address the effects of retrieval practice. Rather, it examined the presence of an FT-like effect when subjects between study of the single passages engaged in semantic generation, a retrieval activity in which subjects generated exemplars from semantic categories unrelated to the study passages (e.g., "Sports" or "Professions"; Divis & Benjamin, 2014, Experiment 2); results indeed showed an FT-like effect in response to semantic generation relative to a distractor baseline condition.

The FT effect with related complex material has been observed in at least four studies to date. Three of the studies divided a coherent video lecture on a single topic (i.e., "Introduction to Statistics" or "Public Health in the 20th Century") into several integrable sections and found an FT effect for the critical final section when subjects engaged in interim retrieval practice between the single sections, both relative to restudy and distractor baseline conditions (Jing et al., 2016; Szpunar et al., 2013, 2014). Similarly, Wissman et al. (2011) reported an FT effect, relative to a no-retrieval-practice condition, in the learning of prose passages that were directly related to one another and contained information intended to be integrated (e.g., on "Capturing and Storing Atmospheric Greenhouse Gases" or "Inconsistencies between Hollywood`s Depiction of History and Factual History"). In this study, recall levels were also compared to a condition in which only the critical final passage was presented for study, with recall in this condition being lower than recall in the retrieval-practice condition, which indicates proactive facilitation. No study has yet examined whether semantic generation can produce an FT-like effect also with related complex material.

Accounts of the FT Effect and the Potential Role of Relatedness of Study Materials

Accounts of the FT effect with complex materials are often similar to those with simple word lists. For instance, for both types of materials, the FT effect has been suggested to reflect release from proactive interference, with retrieval practice insulating the new material against buildup of proactive interference from the preceding materials via an enhancement in "list" discrimination (Divis & Benjamin, 2014; Pastötter et al., 2011; Szpunar et al., 2008). Similarly, for both types of materials, the effect has been attributed to strategy change, assuming that, in response to the experience of retrieval failures during retrieval practice, participants may switch to more efficient strategies for encoding and retrieval of the subsequent material (Chan, Manley et al., 2018; Soderstrom & Bjork, 2014; Wissman et al., 2011). For complex materials, retrieval practice has also been suggested to enhance accessibility of the preceding materials and

facilitate comprehension of the new material, which, in contrast to release-from-proactive-interference, may explain why recall in the retrieval-practice condition can be higher than in a condition in which the final material is studied only (Jing et al., 2016; Wissman et al., 2011). Because semantic generation can trigger mental context change and thus induce "list" discrimination (Jang & Huber, 2008; Shiffrin, 1970), on the basis of the release-from-proactive-interference account, the FT effect should generalize to an FT-like effect in response to semantic generation. No such generalization should arise if the FT effect was mediated by strategy change or facilitated comprehension, both of which predict a retrieval-practice-specific FT effect.

Research on the FT effect with more complex study material is still fragmentary and shows a number of empirical gaps, which is why definitive answers on exactly which mechanisms mediate the FT effect are still missing. One of the critical gaps is the question whether relatedness of study materials influences the FT effect. The issue has recently been addressed in word-list studies with evidence that relatedness of study materials can influence the FT effect. Employing lists of unrelated words as well as lists of words in which different exemplars from the same categories were provided both within and across lists, previous studies have found the FT effect to be present with both types of word lists (e.g., Szpunar et al., 2008). However, only with unrelated lists did FT-like effects arise in response to semantic generation (Kliegl & Bäuml, 2021, 2023). Relatedness of materials thus may also influence the FT effect with more complex study materials, and, for instance, release-from-proactive-interference mediate the effect with unrelated materials and strategy change or facilitated comprehension with highly related materials.

The Present Study

Here, the results of three experiments are reported designed to examine the FT effect with both related and unrelated prose passages. In Experiments 1 and 2, participants studied

five prose passages and, immediately after study of passage 5, were asked to freely recall the critical final passage. Experiment 1 employed passages that were unrelated to each other and covered different topics; Experiment 2 employed passages that were directly related to each other, covered a common topic, and were intended to be integrated. In all other aspects, the two experiments were identical. After study of passages 1 through 4, participants were either asked to recall the immediately preceding passage (retrieval-practice condition), study the immediately preceding passage once again (restudy condition), generate as many exemplars as possible from given semantic categories (semantic-generation condition), or solve simple arithmetic tasks (arithmetic-tasks condition). Both experiments also contained a condition in which only the critical passage 5 was studied. Using similar procedure as Experiments 1 and 2, Experiment 3 examined the effect of relatedness of study materials within a single experiment, including retrieval-practice and restudy conditions as well as the condition in which the final material was studied only.

Three research questions were addressed. First, can interim retrieval practice promote new learning both when subjects study related materials and when they study unrelated materials, thus addressing the issue of whether relatedness of materials can influence the presence of the FT effect? Second, are recall levels for the critical final passage higher after encoding and retrieval of the preceding passages than in the condition in which the final passage is studied only? If so, does this hold regardless of relatedness of study material? And third, do FT-like effects arise if, between the passages, subjects engage in retrieval of information not related to the single passages? Engaging in such retrieval - for instance, by setting up the grocery list for the evening –, is a quite typical way of how students spend breaks between the study of materials, and it is thus important to know whether such retrieval can also benefit new learning and induce effects that mimic those of retrieval practice. We expected FT effects for both the related and the unrelated prose passages. If release-from-proactive-interference

mediated the FT effect with unrelated material and strategy change or facilitated comprehension with related material, the effect should generalize to semantic generation with unrelated passages only and recall in the retrieval-practice condition be higher than in the final-passage-only condition with related passages only.

Experiment 1

Method

Ethical Considerations

All reported experiments were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. Each participant in Experiments 1, 2 and 3 provided informed consent prior to participation.

Participants

140 students of different German universities took part in Experiment 1 (mean age = 23.5 years, range = 18-33 years, 118 female, 22 male), with 28 subjects in each of the five experimental conditions. Sample size was determined on the basis of a power analysis (Faul et al., 2007) setting α to .05 and β to .20 and using the meta-analytic effect size of $g = 0.80$ for the FT effect with prose materials (Chan, Meissner et al., 2018). All participants spoke German as their native language and received monetary reward or course credit for participation. Participants were tested individually in an online video conference via Zoom.

Materials

Five prose passages that were unrelated to one another and covered different topics were employed as study material (all materials and data from the single experiments are available on the Open Science Framework [osf.io/dg5n3]). Each passage was approximately 140 words in length. Four of the passages were already used in prior studies and translated into German: the neandertals (Fritz & Morris, 2015; passage 1), the porcupines (Divis & Benjamin, 2014;

passage 2), the chronic wasting disease (Divis & Benjamin, 2014; passage 4), and the history and uses of garlic (Fritz & Morris, 2015; passage 5). One passage described the story of Johann Beringer`s `lying-stones´ (passage 3) and was based on information in Taylor (2017). Presentation order of the five prose passages was the same for all participants. Thus, passage 5 always served as the critical prose passage, whereas passages 1–4 always served as the prose passages preceding the critical passage.

Design

Four of the five experimental conditions followed a 2 x 2 design with the between-subjects factors of PRACTICE (present vs. absent) and RETRIEVAL (present vs. absent). Practice was present in the restudy and retrieval-practice conditions, in which immediately after study of passages 1–4, participants practiced the preceding passage by virtue of restudy or retrieval practice; practice was absent in the arithmetic-tasks and semantic-generation conditions, in which immediately after study of passages 1–4, participants engaged in arithmetic or semantic-generation tasks that did not require practice of the preceding passages. Retrieval was present in the retrieval-practice and semantic-generation conditions, and absent in the restudy and arithmetic-tasks conditions.

Besides these four conditions in which all five passages were studied in succession, an additional passage-5-only condition was included in the experiment in which passages 1–4 were not presented prior to study of the critical passage 5. Recall in this condition served as a baseline to measure the influence of encoding and retrieval of the preceding passages on critical-passage recall relative to a condition in which no such preceding passages were present.

Procedure

Figure 1 depicts the experiment`s procedure and conditions. In the restudy, retrieval-practice, arithmetic-tasks and semantic-generation conditions, participants studied five prose passages in anticipation of a final cumulative recall test. Prior to the start of the experiment,

participants were told to expect various activities that may follow the presentation of each single passage, which can include restudy of the passage that they had just studied, free recall of the passage that they had just studied, solving simple arithmetic tasks, or generating exemplars from semantic categories unrelated to the prose passages. We pretended that activities following each passage were determined on a random basis. In fact, the activities differed between conditions and participants passed through the same activities after study of passages 1–4 within each experimental condition (cf. Szpunar et al., 2008). All five prose passages were presented in succession for 3 min each at the center of a computer screen. After the presentation of each single passage, participants counted backward in steps of threes from a three-digit number for 30 s. The type of activity that followed the backward counting after passages 1–4 was conducted for 3 min each and varied with experimental condition. Participants were either asked to (i) study the immediately preceding passage once again, which was reexposed on the computer screen (restudy condition), (ii) type in all the information they could remember from the immediately preceding passage in a response box on the computer screen (retrieval-practice condition), (iii) solve simple arithmetic tasks (arithmetic-tasks condition), or (iv) type in as many words as possible from a single semantic category (semantic-generation condition). In the semantic-generation task, subjects after study of each passage were first given 90 s to generate exemplars from one of eight categories (BODY PARTS, FOOD, FOUR-LEGGED ANIMALS, MALE NAMES, MEANS OF TRANSPORT, MUSICAL INSTRUMENTS, OFFICE SUPPLIES, or SPORTS), and then were given another 90 s to generate exemplars from another category. Assignment of the eight categories to activities following passages 1–4 was random for each participant. After study of passage 5 and the backward-counting task, participants in all four conditions were given 3 min to type in as much information as they could remember from passage 5 in a response box on the computer screen. Following passage-5 recall and another 3-min interval filled with Tetris as a distractor task, participants underwent a final cumulative test in which

they had 15 min to type in all the information they could remember from all five passages they had previously studied (performance on this final test, as well as performance during retrieval practice are reported in the Supplementary Material, for archival purposes).

In the passage-5-only condition, passages 1 through 4 were not presented. Rather, participants in this condition were asked to just study passage 5 in anticipation of a final recall test. Presentation of passage 5, subsequent backward counting, and recall of passage 5 followed the procedure in the other four conditions.

Scoring

To assess performance on the passage-5 recall test, passage 5 was divided into 11 idea units. Participants` recall responses were scored by giving 1 point for a correctly recalled idea unit or 0.5 points for a correctly recalled critical part of an idea unit. For example, the idea unit "Sanskrit writings from 5000 years ago refer to garlic and 3000-year-old Chinese manuscripts mention garlic" was scored 1 point for the response "Sanskrit writings from 5000 years ago and Chinese manuscripts from 3000 years ago mention garlic", as the response correctly reproduced the entire idea unit. However, the idea unit was scored 0.5 points for the response "Sanskrit writings from 5000 years ago mention garlic", or for the response "Chinese manuscripts from 3000 years ago mention garlic", as these responses correctly reproduced only parts of the idea unit. Credit was assigned for both verbatim responses or correct paraphrases of idea units from passage 5. Initially, 70 recall protocols were scored by two raters, and the Pearson correlation (*r*) between their scores was .99. Given the high interrater reliability, the remaining passage-5 recall protocols were scored by one rater only.

Results

Correct Passage-5 Recall

Figure 2A shows the percentage of correctly recalled passage-5 idea units for each of the five experimental conditions. Regarding the four conditions in which all five passages were presented for study, a 2 x 2 ANOVA with the between-subjects factors of PRACTICE (present vs. absent) and RETRIEVAL (present vs. absent) revealed no significant main effect of PRACTICE, $F(1, 108) = 0.082$, $MSE = 0.03$, $p = .775$, $\eta^2 < .01$, $B_{01} = 10.191$, but a main effect of RETRIEVAL, $F(1, 108) = 74.431$, $MSE = 0.03$, $p < .001$, $\eta^2 = .41$, reflecting that recall was higher when retrieval was present than when it was absent (60.1% vs. 33.1%). Critically, there was no significant interaction between the two factors, $F(1, 108) = 1.614$, $MSE = 0.03$, $p = .207$, $\eta^2 = .02$, $B_{01} = 4.639$, suggesting that practice did not influence the effect of retrieval. Consistently, planned comparisons showed a reliable difference in recall between the retrieval-practice and restudy conditions (61.7% vs. 30.7%), $t(54) = 6.745$, $p < .001$, $d = 1.80$, 95% CI of the difference = [21.8, 40.2], as well as between the semantic-generation and arithmetic-tasks conditions (58.6% vs. 35.6%), $t(54) = 5.413$, $p < .001$, $d = 1.45$, 95% CI of the difference = [14.5, 31.6], thus demonstrating that retrieval practice induced an FT effect and semantic generation increased recall of passage 5 to a similar degree.

Additional analyses showed that recall in the passage-5-only baseline condition was similar to recall in the retrieval-practice and semantic-generation conditions (passage-5-only vs. retrieval-practice condition: 61.8% vs. 61.7%, $t(54) = 0.039$, $p_{adj} = .969$, $d = 0.01$, 95% CI of the difference = [-8.0, 8.3], $B_{01} = 7.478$; passage-5-only vs. semantic-generation condition: 61.8% vs. 58.6%, $t(54) = 0.726$, $p_{adj} = .942$, $d = 0.19$, 95% CI of the difference = [-5.7, 12.2], $B_{01} = 5.667$), but higher than in the restudy and arithmetic-tasks conditions (passage-5-only vs. restudy condition: 61.8% vs. 30.7%, $t(54) = 7.163$, $p_{adj} = .003$, $d = 1.91$, 95% CI of the difference = [22.4, 39.9]; passage-5-only vs. arithmetic-tasks condition: 61.8% vs. 35.6%, $t(54)$

= 7.408, p_{adj} = .004, d = 1.98, 95% CI of the difference = [19.2, 33.4]). To control the familywise error rate across the four comparisons, p-values were adjusted by employing the sequential Bonferroni procedure.

Intrusions during Passage-5 Recall

Intrusions were defined as idea units from passages 1, 2, 3, or 4 that participants produced during the passage-5 recall test. Critically, not a single participant had an intrusion from a preceding prose passage, irrespective of experimental condition (see Table 1).

Discussion

Using unrelated prose passages as study material, the FT effect was present in response to retrieval practice and an FT-like effect was present in response to semantic generation. These results are consistent with Divis and Benjamin`s (2014, Experiment 2) finding on the effects of semantic generation and extend on Wissman et al.'s (2011, Experiment 1A) and Yue et al.`s (2015, Experiment 2) work by demonstrating that, also with a larger number of unrelated passages, retrieval practice can induce an FT effect relative to a restudy condition. Recall in the passage-5-only condition was similar to recall in the retrieval-practice condition, and the results for the semantic-generation condition, when compared to those in the arithmetic-tasks and passage-5-only conditions, mimicked those for the retrieval-practice condition. These findings are in line with the release-from-proactive-interference account. Intrusions during recall of the critical final passage were absent irrespective of condition. Possibly, intrusions from the preceding passages came to mind but could quickly be identified as incorrect candidates stemming from the (unrelated) preceding passages, and thus were not reported in the recall protocols (see Pierce et al., 2017). The goal of Experiment 2 was to examine whether the findings of Experiment 1 with unrelated prose passages generalize to related prose passages.

Experiment 2

Method

Participants

Another 140 students of different German universities took part in Experiment 2 (mean age = 23.8 years, range = 18-35 years, 115 female, 25 male), with 28 subjects in each of the five experimental conditions. Sample size followed Experiment 1. All participants spoke German as their native language and received monetary reward or course credit for participation.

Materials, Design, and Procedure

Experimental details were identical to Experiment 1, with the only exception of study material. In Experiment 2, we employed five prose passages that were directly related to one another, covered a common topic, and contained information that was intended to be integrated. Specifically, the five passages were consecutive sections of a prose passage about the big bang theory, which was taken from prior wok (Chan et al., 2006) and translated into German (see also Bäuml & Schlichting, 2014, Experiment 2; Wallner & Bäuml, 2017, Experiment 4). As in Experiment 1, each passage was approximately 140 words in length. Again, presentation order was the same for all participants, with passage 5 always serving as the critical prose passage and passages 1–4 always serving as the preceding passages.

Scoring

For scoring of the passage-5 recall responses, passage 5 was divided into 12 idea units. Again, responses were scored by giving 1 point for a correctly recalled idea unit or 0.5 points for a correctly recalled critical part of an idea unit. For example, the idea unit “By late 1989 NASA’s Cosmic Background Explorer was ready for action” was scored 1 point for the response “In 1989 NASA’s Cosmic Background Explorer was ready for action”, as the response

correctly reproduced the entire idea unit. However, the idea unit was scored 0.5 points for the response "In 1989 NASA's satellite was ready for action", or for the response "NASA's Cosmic Background Explorer was ready for action", as these responses correctly reproduced only parts of the idea unit. Like in Experiment 1, credit was assigned for both verbatim responses or correct paraphrases of idea units from passage 5. Again, 70 recall protocols were scored by two raters, and an interrater reliability of $r = .99$ was obtained. Given the high interrater agreement, the remaining passage-5 recall protocols were scored by one rater only.

Results

Correct Passage-5 Recall

Figure 2B shows the percentage of correctly recalled passage-5 idea units for each of the five experimental conditions. Regarding the four conditions in which all five passages were studied, a 2 x 2 ANOVA with the between-subjects factors of PRACTICE (present vs. absent) and RETRIEVAL (present vs. absent) revealed a main effect of PRACTICE, $F(1, 108) = 32.058$, $MSE = 0.03$, $p < .001$, $\eta^2 = .23$, but no significant main effect of RETRIEVAL, $F(1, 108) = 3.221$, $MSE = 0.03$, $p = .075$, $\eta^2 = .03$, $B_{01} = 2.035$, reflecting that, in general, recall was higher when practice was present than when it was absent (37.9% vs. 19.8%), whereas no such difference arose between the presence and absence of retrieval (31.7% vs. 26.0%). Critically, there was also a significant interaction between the two factors, $F(1, 108) = 31.591$, $MSE = 0.03$, $p < .001$, $\eta^2 = .23$, suggesting that the effect of retrieval depended on practice. Indeed, while planned comparisons showed that recall in the retrieval-practice condition was higher than in the restudy condition (49.7% vs. 26.0%), $t(54) = 4.575$, $p < .001$, $d = 1.22$, 95% CI of the difference = [13.3, 34.0], recall in the semantic-generation condition was even lower than in the arithmetic-tasks condition (13.7% vs. 25.9%), $t(54) = 3.266$, $p = .002$, $d = 0.87$, 95% CI of the difference = [4.7, 19.7]. Retrieval practice thus induced an FT effect, but semantic generation did not induce an FT-like effect.

Additional analyses showed that recall in the passage-5-only condition was similar to recall in the restudy and arithmetic-tasks conditions (passage-5-only vs. restudy condition: 24.7% vs. 26.0%, $t(54) = 0.300$, $p_{adj} > .999$, $d = 0.08$, 95% CI of the difference = [-10.2, 7.5], $B_{01} = 7.204$; passage-5-only vs. arithmetic-tasks condition: 24.7% vs. 25.9%, $t(54) = 0.272$, $p_{adj} = .786$, $d = 0.07$, 95% CI of the difference = [-10.0, 7.6], $B_{01} = 7.200$), but was lower than in the retrieval-practice condition (24.7% vs. 49.7%), $t(54) = 4.928$, $p_{adj} = .004$, $d = 1.32$, 95% CI of the difference = [14.8, 35.2] and higher than in the semantic-generation condition (24.7% vs. 13.7%), $t(54) = 3.030$, $p_{adj} = .012$, $d = 0.81$, 95% CI of the difference = [3.7, 18.3]. Like in Experiment 1, the *p*-values for all four comparisons were adjusted by using the sequential Bonferroni procedure.

Intrusions during Passage-5 Recall

Table 1 shows the number of intrusions during passage-5 recall in the restudy, retrieval-practice, arithmetic-tasks, and semantic-generation conditions. A 2 x 2 ANOVA with the factors of PRACTICE and RETRIEVAL revealed no significant main effect of PRACTICE, $F(1, 108) = 1.200$, $MSE = 0.60$, $p = .276$, $\eta^2 = .01$, $B_{01} = 5.702$, but a main effect of RETRIEVAL, $F(1, 108) = 9.257$, $MSE = 0.60$, $p = .003$, $\eta^2 = .08$, reflecting that participants produced fewer intrusions in the presence than the absence of retrieval (0.13 vs. 0.57). There was also a marginally significant interaction between PRACTICE and RETRIEVAL, $F(1, 108) = 3.332$, $MSE = 0.60$, $p = .071$, $\eta^2 = .03$, $B_{01} = 1.930$. Planned comparisons showed a reliable difference in intrusions between the retrieval-practice and restudy conditions (0.07 vs. 0.79), $t(54) = 2.942$, $p = .005$, $d = 0.79$, 95% CI of the difference = [0.23, 1.20], but no reliable difference in intrusions between the semantic-generation and arithmetic-tasks conditions (0.18 vs. 0.36), $t(54) = 1.083$, $p = .283$, $d = 0.29$, 95% CI of the difference = [-0.51, 0.15], $B_{01} = 4.100$.

Discussion

Like in Experiment 1, the FT effect was present and recall higher in the retrieval-practice than the restudy and arithmetic-tasks conditions, which is consistent with prior work with related complex material (Jing et al., 2016; Szpunar et al., 2013, 2014; Wissman et al., 2011). Unlike in Experiment 1, however, recall in the retrieval-practice condition was higher than recall in the passage-5-only condition, and the results in the semantic-generation condition did not mimic those in the retrieval-practice condition. Indeed, correct recall in the semantic-generation condition was even lower than in the arithmetic-tasks and passage-5-only conditions, and, although intrusions were lower in the retrieval-practice than the restudy condition (see also Szpunar et al., 2013), intrusions were not reduced in the semantic-generation relative to the arithmetic-tasks condition. These findings disagree with the release-from-proactive-interference view and rather point to strategy change or facilitated comprehension as possible mechanisms underlying the FT effect.

Experiments 1 and 2 differed considerably in recall levels in the passage-5-only condition, which points to major differences in recall difficulty between the two experiments' final materials. To exclude that the observed differences in results between the two experiments were driven by differences in passage-5 recall difficulty, Experiment 3 manipulated relatedness of materials within a single experiment, comparing recall between a retrieval-practice condition, a restudy condition, and a condition in which passage 5 was studied only.

Experiment 3

Method

Participants

Another 140 students of different German universities took part in Experiment 3 (mean age = 24.1 years, range = 18-34 years, 106 female, 33 male, 1 diverse), with 28 subjects in each

of the five experimental conditions. Sample size followed Experiments 1 and 2. All participants spoke German as their native language and received monetary reward or course credit for participation.

Materials

Like in Experiments 1 and 2, five prose passages were employed as study material. In all experimental conditions, passage 5 used in Experiment 2 ("big bang theory") served as the critical final passage, whereas the preceding passages 1–4 varied between conditions. In the unrelated condition, passages 1–4 used in Experiment 1 ("neandertals", "porcupines", "lying-stones", and "chronic wasting disease") were employed as the preceding passages; in the related condition, passages 1–4 used in Experiment 2 ("big bang theory") were employed. Again, presentation order of passages was held constant for all participants.

Design

Four of the five conditions followed a 2 x 2 design with the between-subjects factors of TYPE OF PRACTICE (restudy vs. retrieval practice) and RELATEDNESS OF MATERIALS (unrelated vs. related). In the unrelated condition, participants were presented with five passages that were unrelated to each other and covered different topics, whereas in the related condition, five passages that were directly related to one another and covered the same topic were presented. Participants were asked after study of passages 1–4 to practice the just presented passage by virtue of restudy or retrieval practice. Again, the experiment included an additional passage-5-only condition in which no preceding passages were presented prior to study of the critical passage 5.

Procedure

The procedure differed in two aspects from Experiments 1 and 2. First, in order to avoid floor effects, time for both study and practice were increased to 4 min per passage. Second, there was no final cumulative test at the end of the single conditions.

Scoring

Scoring of the passage-5 recall responses was conducted following the same criteria that had been used for the scoring procedure in Experiment 2. Again, two independent raters scored 70 of the recall protocols, and an interrater reliability of $r = .99$ was obtained. Given the high interrater agreement, the remaining passage-5 recall protocols were scored by one rater only.

Results

Correct Passage-5 Recall

Figure 3 shows the percentage of correctly recalled passage-5 idea units for each of the five experimental conditions. Regarding the four conditions in which five passages were presented for study, a 2 x 2 ANOVA with the between-subjects factors of TYPE OF PRACTICE (restudy vs. retrieval practice) and RELATEDNESS OF MATERIALS (unrelated vs. related) revealed main effects of TYPE OF PRACTICE, $F(1, 108) = 38.178$, $MSE = 241.60$, $p < .001$, $\eta^2 = .26$, and RELATEDNESS OF MATERIALS, $F(1, 108) = 29.391$, $MSE = 241.60$, $p < .001$, $\eta^2 = .21$, reflecting higher recall in the retrieval-practice than the restudy condition (51.3% vs. 33.1%), and higher recall in the related than the unrelated condition (50.2% vs. 34.2%). Critically, there was no significant interaction between the two factors, $F(1, 108) = 1.240$, $MSE = 241.60$, $p = .268$, $\eta^2 = .01$, $B_{01} = 5.583$, indicating that the FT effect did not depend on relatedness of prose materials. Consistently, planned comparisons between the retrieval-practice and restudy conditions showed reliable FT effects with both the related and unrelated materials (related: 60.9% vs. 39.4%, $t(54) = 4.943$, $p < .001$, $d = 1.32$, 95% CI of the difference = [12.7, 30.1];

unrelated: 41.7% vs. 26.8%, $t(54) = 3.751$, $p < .001$, $d = 1.00$, 95% CI of the difference = [6.9, 22.8]). For both the restudy and retrieval-practice conditions, recall was higher after study and practice of the related than the unrelated materials (restudy: 39.4% vs. 26.8%, $t(54) = 3.264$, $p = .002$, $d = 0.87$, 95% CI of the difference = [4.8, 20.4]; retrieval practice: 60.9% vs. 41.7%, $t(54) = 4.348$, $p < .001$, $d = 1.16$, 95% CI of the difference = [10.3, 28.0]).

Additional analyses showed that, for unrelated materials, recall in the passage-5-only condition was similar to recall in the retrieval-practice condition (40.6% vs. 41.7%), $t(54) = 0.244$, $p_{adj} = .808$, $d = 0.07$, 95% CI of the difference = [-9.6, 7.5], $B_{01} = 7.256$, but higher than in the restudy condition (40.6% vs. 26.8%), $t(54) = 3.422$, $p_{adj} = .003$, $d = 0.92$, 95% CI of the difference = [5.7, 21.9]. For related materials, in contrast, recall in the passage-5-only condition was similar to recall in the restudy condition (40.6% vs. 39.4%), $t(54) = 0.284$, $p_{adj} > .999$, $d = 0.08$, 95% CI of the difference = [-7.2, 9.6], $B_{01} = 7.178$, but lower than in the retrieval-practice condition (40.6% vs. 60.9%), $t(54) = 4.512$, $p_{adj} = .004$, $d = 1.21$, 95% CI of the difference = [-29.2, -11.2]. For all four comparisons to the passage-5-only condition, *p*-values were again adjusted following the sequential Bonferroni procedure. The results from the analyses mirror those reported in the retrieval-practice, restudy, and passage-5-only conditions of Experiments 1 and 2.

Intrusions during Passage-5 Recall

Table 1 shows the number of intrusions during passage-5 recall in the restudy and retrieval-practice conditions, both with related and with unrelated passages 1–4. With unrelated passages, none of the participants had an intrusion from a preceding passage, irrespective of practice condition. With related passages, in contrast, intrusion rates were significantly greater than zero ($M = 0.52$), $t(55) = 4.334$, $p < .001$, $d = 0.58$, 95% CI of the difference = [0.28, 0.76], and lower in the retrieval-practice than the restudy condition (0.18 vs. 0.86), $t(54) = 3.045$, $p =$

.004, d = 0.81, 95% CI of the difference = [0.23, 1.13]. Again, results mirror those reported in the retrieval-practice and restudy conditions of Experiments 1 and 2.

General Discussion

This study demonstrates that retrieval practice can produce an FT effect with both related and unrelated prose passages. In fact, for both types of passages, retrieval practice on the previously encoded materials created higher recall levels on the critical final material than restudy of the preceding materials or a distractor task did (Experiments 1-3). At the same time, two differences in findings emerged between the two types of material. The one difference was that the FT effect generalized to an FT-like effect in response to semantic generation with unrelated prose passages but not with related passages (Experiments 1 and 2). The second difference was that retrieval practice induced a recall level for the final critical passage that was similar to the recall level in the passage-5-only condition with unrelated prose materials (Experiments 1 and 3) but above this level with related prose materials (Experiments 2 and 3).

The differences in findings between related and unrelated study materials suggest that, despite the presence of the same basic FT effect for the two types of materials, different cognitive mechanisms may have contributed to the FT effect in the two situations. The results for the unrelated passages revealed an equivalence of effects of retrieval practice and semantic generation and a recall level in the retrieval-practice condition that was similar to the level in the passage-5-only condition, which is consistent with the release-from-proactive-interference account, according to which the FT effect reflects an enhancement in "list" discrimination. In contrast, the results for the related passages showed a difference in results between retrieval practice and semantic generation and a recall level in the retrieval-practice condition that exceeded that in the passage-5-only condition, which is consistent with the strategy-change and facilitated-comprehension accounts of the FT effect, both of which suggest retrieval-practice-specificity of the FT effect and are open to effects of proactive facilitation. In line with prior

word-list studies (Kliegl & Bäuml, 2021, 2023), the results thus provide evidence that qualitatively different FT effects can arise for related and unrelated complex material.

This study employed prose passages that covered different topics and were unrelated to each other as well as prose passages that covered the same topic, were directly related to one another, and were intended to be integrated. On a hypothetical relatedness scale, the two conditions thus reflect quite extreme points, raising the question of how results for the FT effect might look like for intermediate levels of relatedness. Wissman et al. (2011, Experiment 1A) employed such intermediate level using texts that shared a common topic but were otherwise not directly related to one another and had no overlapping information. Results showed an FT effect relative to a no-retrieval-practice condition, but the study did not include a semantic-generation or a final-passage-only condition, so that a more detailed comparison of results with those from the present study is not possible. Still, Wissman et al. (p. 1146) raised doubt on whether topically related but otherwise separate texts would lead to similar results than highly related texts, suggesting that, for such intermediate relatedness levels, the roles of facilitated comprehension or strategy change might be reduced. Clarifying the exact roles of release-from-proactive-interference and strategy change/facilitated comprehension for the FT effect with intermediate relatedness levels in fact is a high priority for future work.

The present findings are potentially important for educational practice and provide some useful clues with regard to the putative courses scenario entertained above. First, retrieval practice can benefit learning and memory of future material both when highly related and when unrelated materials are studied, suggesting that, regardless of whether students, during exam preparation, are facing the study of highly related materials from the same course or quite unrelated materials from different courses, inbetween-retrieval practice can optimize new learning. Second, with unrelated materials, future learning can also benefit from retrieval of other information than the just studied material, so that setting up the evening grocery list

between study of the single materials may also improve learning of forthcoming materials. However, and in contrast to interim retrieval practice, doing so may come with costs and not improve memory for the preceding materials (see Divis & Benjamin, 2014). Finally, when students are facing the study of highly related materials, it does not pay off to skip the preceding materials, as in such case, learning and memory for the final materials may be much worse than when all preceding materials were attended and retrieval practiced. Per se this finding may not sound very surprising, but the size of the benefit of encoding and retrieving the preceding material may well exceed a priori expectations.

Conclusions

The results of this study show that retrieval practice can promote new learning with both related and unrelated prose materials, This basic FT effect generalized to an FT-like effect in response to semantic generation with unrelated but not related prose materials, and only with related prose materials did the FT effect lead to recall levels superior to those when no preceding material was studied. Theoretically, the findings suggest that the cognitive mechanisms mediating the FT effect differ for related and unrelated study materials. Empirically, the findings point to the potential power of retrieval practice to enhance new learning in educational settings.

References

Bäuml, K.-H. T., & Schlichting, A. (2014). Memory retrieval as a self-propagating process. *Cognition, 132*(1), 16–21.

Chan, J. C. K., Manley, K. D., Davis, S. D., & Szpunar, K. K. (2018). Testing potentiates new learning across a retention interval and a lag: a strategy change perspective. *Journal of Memory and Language, 102*, 83–96.

Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135*(4)*,* 553–571.

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.

Divis, K. M., & Benjamin, A. S. (2014). Retrieval speeds context fluctuation: Why semantic generation enhances later learning, but hinders prior learning. *Memory & Cognition, 42*(7), 1049–1062.

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

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Frase, L. T. (1968). Some data concerning the mathemagenic hypothesis. *American Educational Research Journal, 5,* 181–189.

Fritz, C. O., & Morris, P. E. (2015). Partset cuing of texts, scenes, and matrices. *British Journal of Psychology, 106*(1), 1–21.

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.

Jing, H. G., Szpunar, K. K., & Schacter, D. L. (2016). Interpolated testing influences focused attention and improves integration of information during a video-recorded lecture. *Journal of Experimental Psychology: Applied, 22*(3), 305–318.

Kliegl, O., & Bäuml, K.-H. T. (2021). When retrieval practice promotes new learning – the critical role of study material. *Journal of Memory and Language, 120,* 104253.

Kliegl, O., & Bäuml, K.-H. T. (2023). How retrieval practice and semantic generation affect subsequently studied material: An analysis of item-level effects. *Memory, 31,* 127–136.

Kriechbaum, V. M., & Bäuml, K.-H. T. (2024, January 29). Retrieval Practice Can Promote New Learning With Both Related and Unrelated Prose Materials. Retrieved from osf.io/dg5n3

Pastötter, B., & Bäuml, K.-H. T. (2014). Retrieval practice enhances new learning: the forward effect of testing. *Frontiers in Psychology, 5,* 286.

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(2)*,* 287–297.

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.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1)*,* 20–27.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255.

Rothkopf, E. Z. (1966). Learning from written instructive material: an exploration of the control of inspection behavior by test-like events. *American Educational Research Journal, 3,* 241–250.

Shiffrin, R. M. (1970). Forgetting: Trace erosion or retrieval failure? *Science, 168*(3939), 1601–1603.

Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language, 73,* 99–115.

Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition, 3*(3), 161–164.

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America, 110*(16), 6313–6317.

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1392–1399.

Taylor, P. D. (2017). Fake fossils by the hundred – the story of Johann Beringer`s “lying-stones". *Deposits Magazine* (52), 38–41.

Wallner, L., & Bäuml, K.-H. T. (2017). Beneficial effects of selective item repetition on the recall of other items. *Journal of Memory and Language, 95,* 159–172.

Wissman, K. T., & Rawson, K. A. (2015). Grain size of recall practice for lengthy text material: Fragile and mysterious effects on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(2), 439–455.

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*(6), 1140–1147.

Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: A review of the forward testing effect. *npj Science of Learning, 3*, 8.

Yue, C. L., Soderstrom, N. C., & Bjork, E. L. (2015). Partial testing can potentiate learning of tested and untested material from multimedia lessons. *Journal of Educational Psychology, 107*(4), 991–1005.

Figure Captions

Figure 1. Procedure and conditions employed in Experiments 1 and 2. In four experimental conditions, participants studied five prose passages in succession. Each passage was followed by a short distractor task. After study of passages 1 through 4, participants either solved simple arithmetic tasks (arithmetic-tasks condition), generated exemplars from semantic categories unrelated to the passages (semantic-generation condition), restudied the immediately preceding passage (restudy condition), or freely recalled the passage (retrieval-practice condition). Passage 5 was the critical passage and was freely recalled immediately after study. All five passages were tested in a final recall test. In an additional passage-5-only baseline condition, participants studied and freely recalled passage 5 in the absence of the study of any preceding passages.

Figure 2. (A) Results of Experiment 1. Percentage of correctly recalled passage-5 idea units for unrelated prose passages as a function of condition: arithmetic tasks (no practice, no retrieval), semantic generation (no practice, retrieval), restudy (practice, no retrieval), retrieval practice (practice, retrieval). (B) Results of Experiment 2. Percentage of correctly recalled passage-5 idea units for related prose passages as a function of condition: arithmetic tasks (no practice, no retrieval), semantic generation (no practice, retrieval), restudy (practice, no retrieval), retrieval practice (practice, retrieval). In both graphs, the dashed line represents mean recall in the passage-5-only condition. Error bars represent standard errors.

Figure 3. Results of Experiment 3. Percentage of correctly recalled passage-5 idea units as a function of type of practice (restudy, retrieval practice) and relatedness of prose passages (unrelated passages, related passages). The dashed line represents mean recall in the passage-5-only condition. Error bars represent standard errors.

Figure 1

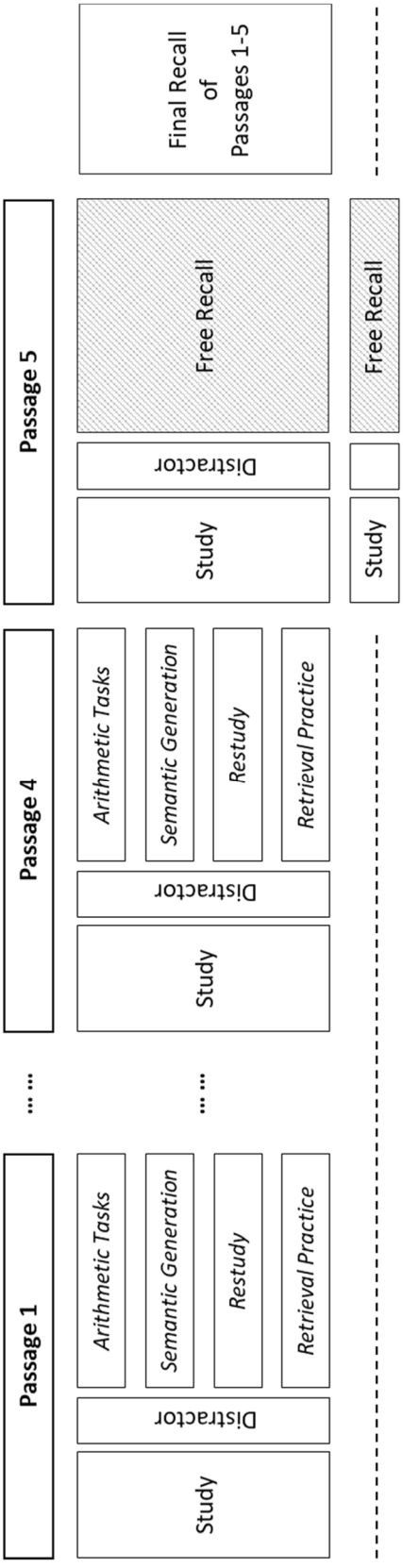
Passage 1
Study
Distractor
Arithmetic Tasks
Semantic Generation
Restudy
Retrieval Practice
... ...
Passage 4
Study
Distractor
Arithmetic Tasks
Semantic Generation
Restudy
Retrieval Practice
... ...
Passage 5
Study
Distractor
Free Recall
Study
Free Recall
Final Recall of Passages 1-5

Figure 2

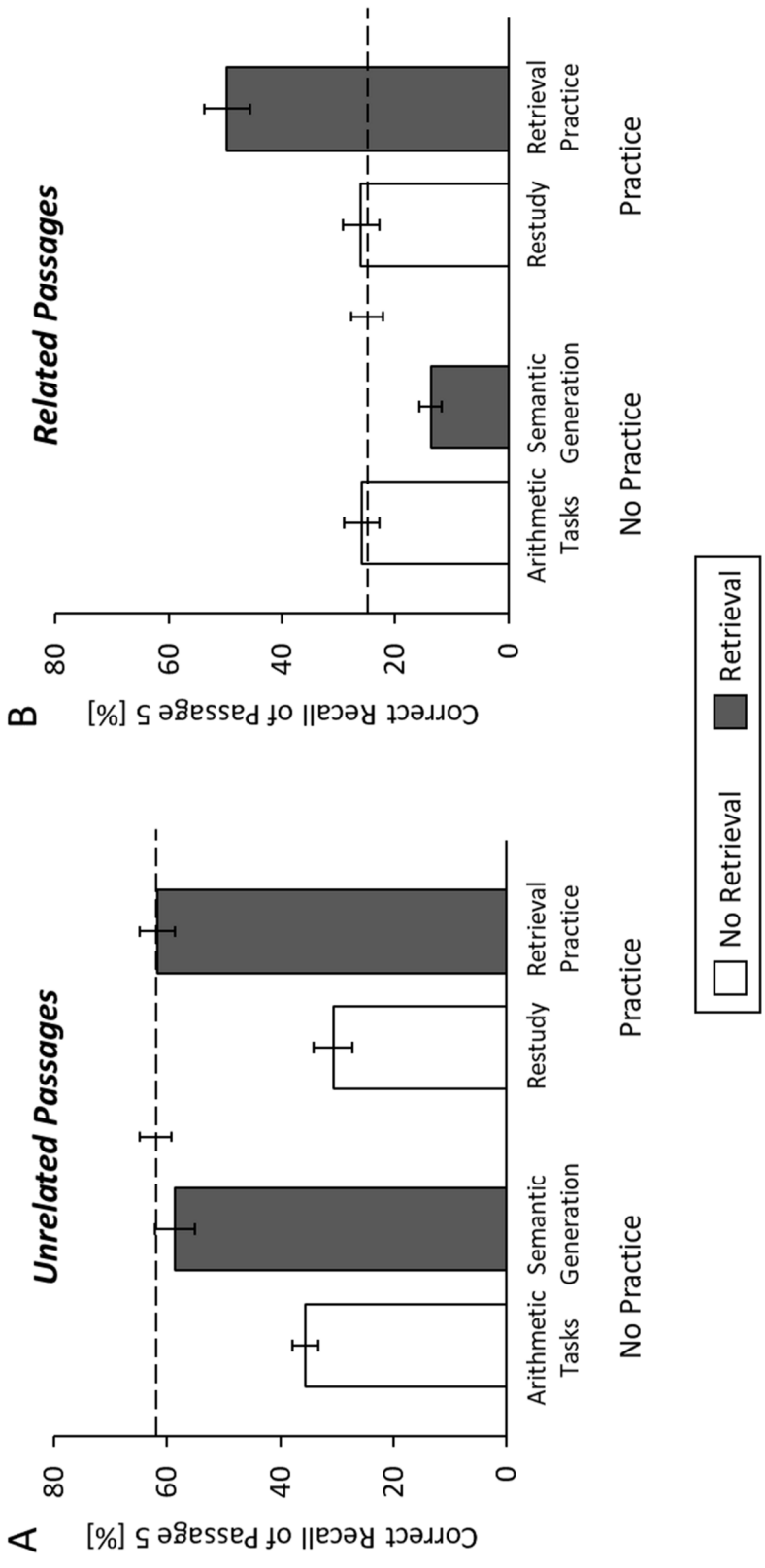
A
Unrelated Passages
Correct Recall of Passage 5 [%]
0
20
40
60
80
Arithmetic Tasks
Semantic Generation
Restudy
Retrieval Practice
No Practice
Practice
B
Related Passages
Correct Recall of Passage 5 [%]
0
20
40
60
80
Arithmetic Tasks
Semantic Generation
Restudy
Retrieval Practice
No Practice
Practice
No Retrieval
Retrieval

Figure 3

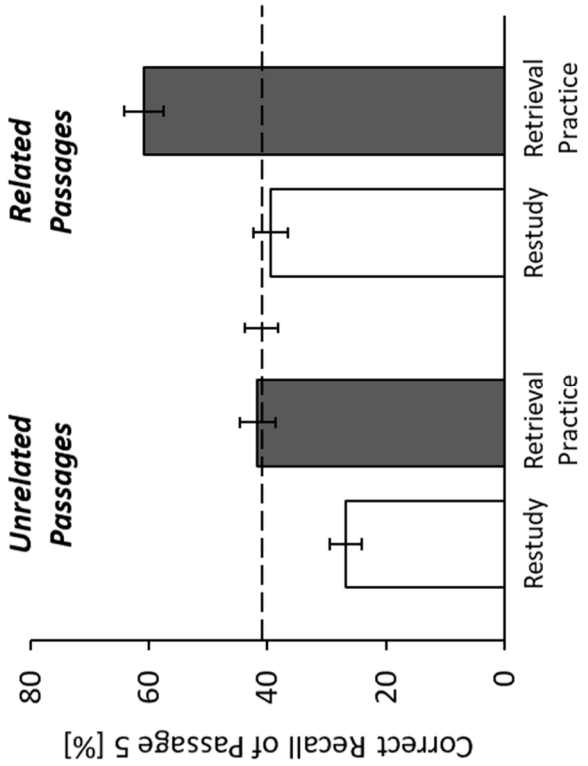
Correct Recall of Passage 5 [%]
0
20
40
60
80
Unrelated Passages
Related Passages
Restudy
Retrieval Practice
Restudy
Retrieval Practice

Table 1.

Mean number of intrusions from prose passages 1–4 during recall of passage-5 idea units for Experiments 1, 2, and 3 (standard errors in parenthesis).

Condition	Arithmetic Tasks	Semantic Generation	Restudy	Retrieval Practice
Experiment 1	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Experiment 2	0.36 (0.14)	0.18 (0.09)	0.79 (0.24)	0.07 (0.05)
Experiment 3				
unrelated passages	---	---	0.00 (0.00)	0.00 (0.00)
related passages	---	---	0.86 (0.20)	0.18 (0.09)