AMERICAN PSYCHOLOGICAL ASSOCIATION
Journal of Experimental Psychology: Applied

ISSN: 1076-898X
https://doi.org/10.1037/xap0000540

Benefits of Pretesting Prior to Retrieval Practice Are Limited, Unless Used for Prior Knowledge–Based Personalization

Thomas Wilschut[1], Maarten van der Velde[1, 2], Florian Sense[3], Bridgid Finn[4], Burcu Arslan[5], and Hedderik van Rijn[1, 2]
[1] Department of Experimental Psychology, University of Groningen
[2] MemoryLab, Groningen, The Netherlands
[3] Infinite Tactics, Beavercreek, Ohio, United States
[4] Educational Testing Service, Princeton, New Jersey, United States
[5] Educational Testing Service Global, Amsterdam, The Netherlands

An abundance of research has demonstrated that both posttesting (also referred to as *retrieval practice*) and pretesting (asking a learner for an answer to a cue *before* a study trial has been provided) can enhance the long-term retention of material. While the benefits of retrieval practice have been widely applied in various real-world applications, such as computerized tools that promote the memorization of factual materials, pretesting has seen limited real-world application. In this study, we examine whether and under which realistic digital learning conditions combining pretesting and posttesting can promote learning. In four experiments (total $N = 210$), we contrast learning conditions in which repeated retrieval practice is preceded by passive study to learning conditions in which retrieval practice is preceded by a test. In the first two experiments, we confirm and extend previous findings by demonstrating that pretesting boosts retrieval accuracy and reduces response times on subsequent retrieval repetitions, regardless of the accuracy of the pretest. We find these effects both when a fixed item repetition schedule is used and with performance-based, adaptive item scheduling that resembles popular digital learning tools. However, after three repetitions of an item, the initial advantage of pretesting disappears, calling into question its usefulness in applied settings that involve spaced repetition. In the final two experiments, we explore a more targeted use of pretesting, leveraging it to assess prior knowledge. Dropping items that were answered correctly during the pretest enhanced overall learning efficiency, especially for learners with moderate to high prior knowledge, without disadvantaging those with low prior knowledge.

Public Significance Statement
The benefits of actively testing oneself as a way of learning new materials have been well established in scientific literature. Accordingly, many popular digital learning applications use repeated tests to enhance learning. At the same time, traditional research on the benefits of testing has typically focused on either testing *before* a passive study session or testing *after* a passive study session. Here, we explore the advantage of combining the two approaches by designing a method in which the learner is quizzed on materials multiple times without presenting the materials for passive study first. In this approach, learning is facilitated by the presentation of feedback after each test question. We show—using realistic methods and materials that make it possible to apply our conclusions in educational settings—that the benefits of replacing an initial study opportunity with an initial test, prior to the following test practice questions, are quite limited in terms of overall learning outcomes. However, if the initial tests are used to identify the amount of prior knowledge a learner has on the materials, and if this information is then used to further personalize the item schedule, we find robust benefits. Importantly, we find that pretesting does not impair learners with limited amounts of prior knowledge. These results can be used to further improve digital learning systems that tailor the learning sessions toward the needs of individual learners.

Keywords: pretesting, prior knowledge, realistic learning, retrieval practice, test-enhanced learning

Carolina Ellen Kuepper-Tetzel served as action editor.
Thomas Wilschut https://orcid.org/0000-0002-1976-6239
The adaptive learning system discussed in this article is licensed to SlimStampen Besloten Vennootschap, a University of Groningen–supported spin-off directed by Hedderik van Rijn, where Maarten van der Velde now works as senior scientist and Thomas Wilschut worked part-time on topics unrelated to this work. No commercial or financial interests have influenced the setup, analysis, or reporting of this study. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationship that could be construed as potential conflicts of interest. This work was supported by a personal grant issued by the University of Groningen, Faculty of Behavioral and Social Sciences, to Thomas Wilschut.
Thomas Wilschut played a lead role in investigation, methodology, software, formal analyses, and writing–original draft and an equal role in conceptualization and writing–review and editing. Maarten van der Velde played a supporting role in formal analysis and an equal role in conceptualization and writing–review and editing. Florian Sense played a supporting role in formal analyses and an equal role in conceptualization, supervision, and writing–review and editing. Bridgid Finn played an equal
continued

Numerous studies have shown that retrieval practice, also called posttesting, can boost learning (e.g., see Agarwal et al., 2021; Carpenter, 2023; Carpenter et al., 2022; Karpicke & Aue, 2015; Karpicke & Blunt, 2011). Actively attempting to recall information from memory has been shown to be particularly effective in enhancing long-term retention (Roediger & Butler, 2011). Meta-analyses by Adesope et al. (2017), Latimier et al. (2021), and Rowland (2014) report robust effect sizes for the benefits of testing relative to control conditions wherein a nontesting activity occurs (e.g., passively reading or studying information). This insight has spurred the design of computer-based tools that promote the memorization of factual information (e.g., vocabulary items) by presenting retrieval practice questions to learners. Popular learning applications like Anki (https://apps.ankiweb.net), Babbel (https://babbel.com), Duolingo (https://www.duolingo.com), MemoryLab (https://memorylab.nl/en), Memrise (https://www.memrise.com/), Rosetta Stone (https://rosettastone.com), and Quizlet (https://quizlet.com/) all use this approach and are collectively used by hundreds of millions of learners around the world.

A number of experiments show that the size of the testing effect increases with repeated retrieval practice attempts for each item (e.g., see Karpicke & Roediger, 2008; Rawson & Dunlosky, 2011). In typical study setups, the expected answer is given along with the cue (*prestudy*) when learners first encounter a new item. This fits with the idea that the initial presentation might also be the learner's first encounter with that specific item. However, both laboratory and educational practice studies have demonstrated that presenting only the cue and asking for a (guessed) response on the initial encounter of an item can be beneficial—even if this *pretesting* results in an incorrect response—as long as the learner receives feedback following the retrieval attempt (Arnold & McDermott, 2013; Izawa, 1970; Kornell et al., 2009; Kornell & Vaughn, 2016; Metcalfe, 2017; Yan et al., 2014). Therefore, these lab studies could be interpreted as evidence in favor of replacing the default *prestudy* trials—that are commonly used in digital learning systems—with *pretesting* trials followed by feedback. In this work, we will consider the benefits of combining pretesting and repeated posttesting in an applied (e-)learning context.

Pretesting, also referred to as errorful generation or prequestioning, involves taking practice tests before studying new information rather than afterward. Traditional studies examining the benefits of pretesting (e.g., see Kornell et al., 2009; Soderstrom & Bjork, 2023) contrasted a prestudy condition, in which both a cue and a semantically weakly related answer were shown at the initial presentation of a word pair (e.g., "frog-pond"), to a pretesting condition (e.g., "frog-?"), in which the participant was asked to attempt to retrieve an answer at the initial presentation of the word pair. Importantly, the weak relation between cue and answer was chosen to ensure that retrieval attempts were mostly unsuccessful (i.e., participants were essentially guessing the answer) and thus to ensure low accuracy on the pretesting trial. In both conditions, the initial trial was then followed by one study trial, in which both the cue and the answer were shown. Recall accuracy on a later test was reliably higher for items that were pretested compared to items that were prestudied (Kornell et al., 2009). To date, pretesting effects have been found to persist over a wide range of materials, including basic facts like foreign language vocabulary (Potts & Shanks, 2014), but extending to more complex materials such as video lectures (Carpenter & Toftness, 2017) and science texts (Richland et al., 2009). Recent work also successfully demonstrated the beneficial effects of pretesting in direct educational contexts, where undergraduate students who participated in a multiple-choice quiz before the start of the course performed better on both pretested and nonpretested materials, an effect explained by increased attentional processing during class and enhanced self-regulated study outside of class (Soderstrom & Bjork, 2023). Pretesting effects have been demonstrated for different retention intervals (e.g., Kornell et al., 2009) and for different response formats (e.g., multiple choice and cued recall; e.g., see Little & Bjork, 2016).

Theoretical accounts of the benefits of pretesting focus mainly on (a) the benefits of generating errors and/or (b) enhanced subsequent processing of information following the pretest trial (Kornell & Vaughn, 2016). Pan and Carpenter (2023) proposed a three-stage framework to explain the mechanisms underlying pretesting. In the first stage, the pretest triggers psychological processes that are not triggered by passive study methods. These processes can be general, affecting the learners' psychological state (e.g., the pretest might enhance the learners' curiosity; Geller et al., 2018), or more question-specific (e.g., the pretest may cause the formation of memories related to the question, or it may cause the generation of a set of possible answers to a question; Vaughn & Rawson, 2012). The second stage is the "learning stage," in which the answer is shown to the participants. The exact way in which the answer comes to mind—either by successful pretesting, by external feedback, or by passively studying the information—is proposed to only marginally impact learning. In the third stage, the learner is presented with a posttest. Pan and Carpenter (2023) assumed that there are three possible ways in which the processes triggered in the pretesting stage could affect posttest performance. First, the altered general psychological state of the learner could improve learning during the second state indirectly. Second, the item-specific processes triggered at pretest might affect learning in the second stage more directly, as the memories formed in the pretesting stage may drive specific learning behavior. Finally, memories formed during pretest might serve as retrieval cues for the posttest, largely bypassing the second stage.[1]

The above studies suggest that including a prestudy trial prior to multiple-repetition retrieval practice—as is common in most computerized learning applications that are used in applied settings—may not be the most optimal strategy, as pretesting seems to benefit retention even when the learner does not know the correct answer. However, there are several factors that make it difficult to directly

[1] As we focus specifically on the application of the pretesting effect in multiple repetition fact learning, we do not cover studies on its theoretical underpinnings or effects in other contexts, such as free text learning, video lectures, or math courses.

role in conceptualization and writing–review and editing. Burcu Arslan played an equal role in conceptualization and writing–review and editing. Hedderik van Rijn played an equal role in conceptualization, formal analysis, supervision, and writing–review and editing.

Correspondence concerning this article should be addressed to Thomas Wilschut, Department of Experimental Psychology, University of Groningen, Grote Kruissstraat 2/1, 9712TS Groningen, The Netherlands. Email: t.j.wilschut@rug.nl

apply the conclusions of existing pretesting studies to the current context. First, although some studies have aimed to *compare* the effects of pretesting and posttesting (e.g., see Latimier et al., 2019; Pan & Sana, 2021), to the best of our knowledge, no work has directly examined the benefits of *combining* the two approaches in a multiple-repetition paradigm. Our approach differs from typical pretesting studies in two key ways. First, instead of presenting a study opportunity after the pretesting trial, we present the learner with feedback on the pretesting trial, followed by posttest trials. The feedback after the pretest trial serves as a learning opportunity and mimics how feedback is provided in real-world retrieval practice settings. Second, following realistic digital learning contexts, we will repeat the posttests multiple times. It seems intuitively plausible that the beneficial effects of the initial pretest are largest on the posttest immediately following that pretest, with reduced benefits on subsequent posttests. If a pretest affects the general psychological state of the learner (e.g., if it boosts the learner's curiosity or motivation to retrieve the answer to a question), that effect is likely to decrease over time. In addition, each of the posttest repetitions might trigger processes that are similar to those triggered by the pretest, decreasing the relative importance of the initial pretest. Furthermore, each of the posttests may lead to the formation of memories related to the question, reducing the relative importance of the memories created at the pretest. Currently, however, it is unclear to what extent the pretesting effect will persist over multiple posttest repetitions as no studies have directly addressed this.

Relatedly, although several studies found that pretesting can enhance learning, they did not show whether, or under what conditions, pretesting results in the most efficient use of study time available. In fact, although Izawa (1970) showed that participants learned materials quicker when the learning was preceded by a larger number of pretest trials (relative to when learning was preceded by only one pretest trial), she found that participants *successfully memorized a higher number of raw items* (i.e., more correct responses on the test) in the single-pretest condition than in the multiple-pretesting condition. Although recent work did find benefits of pretesting in studies wherein time-on-task has been strictly controlled (for a review, see Pan & Carpenter, 2023), studies suggest that pretesting individual items can be conceptualized as a trade-off: Pretesting has a time cost but also results in better learning outcomes. Choosing pretesting over prestudy is only beneficial if the upfront time cost is outweighed by the learning benefits down the line. Although it is likely that this time cost will vary with the type of materials or the strategies of the learner, here we expect that presenting a pretesting trial creates an additional retrieval opportunity that will improve retrieval accuracy on later repetitions of the same item, but that it will also take more time than a self-paced prestudy trial. However, when multiple retrieval attempts follow the initial trial, it seems likely that the faster responses on these attempts eventually nullify the relative initial time cost of starting the learning session with a pretesting rather than a prestudy trial. To date, it is unclear under which exact circumstances (e.g., for what number of retrieval practice trials following the initial trial) the benefit of pretesting persists and outweighs the associated additional time cost.

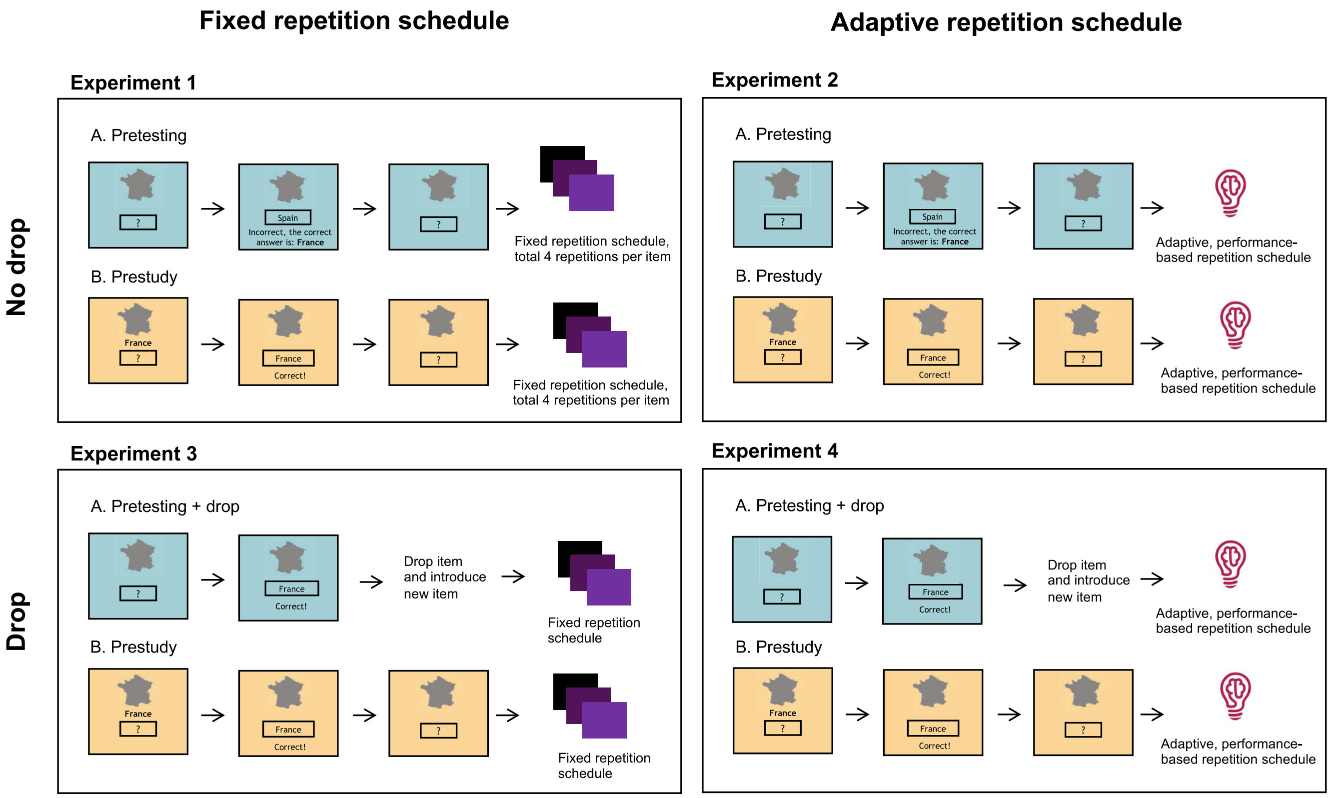
Finally, existing work on pretesting often uses study materials that were specifically selected to ensure that participants have no prior knowledge about them, a quality that is beneficial from an experimental design perspective. However, outside the laboratory, learning rarely occurs without any prior knowledge of the study materials. In fact, the level of prior knowledge is generally considered to be among the most important factors predicting learning outcomes (Brod, 2021; Dochy et al., 2002; Hailikari et al., 2007; Simonsmeier et al., 2022; Thompson & Zamboanga, 2003; Witherby & Carpenter, 2022). Prior knowledge moderates the effectiveness of retrieval practice versus passive study, as seen in so-called expertise-reversal effects (e.g., Kalyuga, 2014). For example, learners with low prior knowledge have been shown to benefit more from studying worked-out examples of problems than solving those same problems, while the reverse has been shown to be true for high prior knowledge learners (Kalyuga et al., 2001). This reversal in strategy effectiveness with increased prior knowledge has been explained in terms of cognitive load (or [working] memory demands; Zambrano et al., 2019): Complex learning tasks may overload the capacity of low prior knowledge learners, who would benefit more from strategies that reduce that load (e.g., studying examples). Considering the effects of prior knowledge on pretesting specifically, earlier studies have shown that semantic relatedness matters for the pretesting effect (e.g., see Grimaldi & Karpicke, 2012). Therefore, it seems plausible that a learner's ability to activate a larger part of the semantic network (e.g., to place the pretested item into a semantic context) could have beneficial effects on pretesting. Yet, we only know of one study that has examined the interaction between pretesting benefits and prior knowledge. Buchin and Mulligan (2023) assigned participants to be trained in multiple topics within one of two academic domains over 3 days. Participants then studied new passages of text related to their trained or untrained domain and completed restudy or retrieval with feedback sessions. Two days later, participants took a final test on the previously learned information from both domains. Importantly, although prior knowledge had a strong effect on test performance (test scores were higher for trained than for untrained knowledge), the benefits of pretesting were nearly identical for high and low prior knowledge information, suggesting that in this case prior knowledge is not a critical boundary condition of retrieval-based learning. However, Buchin and Mulligan (2023) do point out that their results should be replicated with other study materials and tasks, underlining the need to investigate the effects of prior knowledge on the benefits of pretesting in applied settings. Overall, clear evidence on the effects of prior knowledge on the benefits of pretesting in multiple-repetition retrieval practice is lacking.

Current Experiments

In this project, we focused on the effects of pretesting prior to multiple-repetition retrieval practice, which is a frequently used format in digital learning tools. The main goal of this study was to examine whether, and under what conditions, pretesting leads to better learning relative to starting with a prestudy trial. Here, we conducted four experiments to answer this question (see Figure 1 for an outline of the designs). In all experiments, participants learned to link the names of countries to their outlines. Based on our earlier work, we assumed that they were likely to have varying amounts of prior knowledge of this material (Wilschut et al., 2023). Participants completed both a prestudy and a pretesting condition using different study materials. In the prestudy condition, upon the first presentation of an item, the country outline and its correct name were shown. In the pretesting condition, the correct answer was not shown, and the participant was asked to retrieve or guess the correct answer upon the first presentation of the item. In both conditions, the item was

Figure 1

Trial Sequence for the First Presentation of an Item in All Experiments



Note. All four experiments compared a pretesting condition to a prestudy condition. In the pretesting condition, the participant was asked to name the country by its outline on its first presentation. In the prestudy condition, a country outline was shown, together with its written name on the first presentation. In both conditions, the participant was asked to type the retrieved, guessed, or on-screen presented name. Upon the next presentation of the same item, the outline was shown without the name, and the participant was asked to retrieve the country name from the outline. Multiple items were presented, but for visual simplicity, only one item is shown here. In Experiments 1 and 3, items were presented using a fixed item repetitions schedule. In Experiments 2 and 4, items were scheduled using an adaptive, performance-based scheduling system (which we describe in more detail in the main text). In Experiments 3 and 4, items were removed from the practice set upon successful initial retrieval in the pretesting condition. See the online article for the color version of this figure.

then repeated multiple times, and the participant was asked to actively retrieve the item.

In the first experiment, we aimed to quantify the costs and benefits of pretesting over multiple repetitions of an item in a controlled learning session where all items are repeated in a fixed order and for an equal number of times for all learners. To foreshadow the results, the first experiment replicated and extended earlier findings (e.g., Kornell et al., 2009; Metcalfe, 2017) by showing that pretesting boosts subsequent retrieval of the item, independent of the success of the retrieval attempt. However, from the third repetition of an item, the performance difference between pretesting and prestudy disappeared.

In the second experiment, we examined whether these results generalize to adaptive learning contexts in which the item scheduling is not fixed but determined by a learning algorithm developed by *MemoryLab* (see https://memorylab.nl/en) that adapts to the performance of individual learners. As such, the scheduling in this learning condition more closely resembles realistic digital learning, where not all items are introduced in the same fixed order, and some items are repeated more often than other items (e.g., see Mettler et al., 2016; Nakata, 2011). Similar to the fixed-scheduled learning session, we found that pretesting also boosts retrieval performance relative to prestudy in an adaptive learning context. Again, at later repetitions, we found that the benefit of pretesting over prestudy disappeared.

The first two experiments showed that pretesting enhances retrieval performance on subsequent repetitions of the item, but also that this pretesting benefit diminishes over repetitions, calling into question the usefulness of pretesting in applied contexts (i.e., if multiple repetitions are used). In a third experiment, we aimed to explore whether we can use pretesting trials to identify differences in prior knowledge—information that is already known by the learner and therefore does not need to be rehearsed—between learners. We dropped items successfully retrieved on first presentation from further practice and replaced them with new items to maximize learning efficiency. This approach enhanced learning, especially when prior knowledge was high.

Finally, in a fourth experiment, we aimed to extend these results by combining the prior knowledge–based personalization approach from the third experiment with an adaptive-scheduling algorithm.

We found that replacing items that were answered correctly on pretesting trials with new items also resulted in improved learning efficiency in an adaptive learning session.

Experiment 1: Pretesting Benefits in Fixed-Schedule Retrieval Practice

Method

Participants

In total, a group of 104 first-year psychology students studying at the University of Groningen, of whom 79 were female and 25 male, completed either Experiment 1 or 2, which were run during the same testing sessions (see below). The mean age of the participants was 20.3 ($SD = 3.7$) years. Most participants were Dutch or German, but the participant pool also included participants from other European countries. Participants gave informed consent, and the study was approved by the ethical committee of the department of psychology at the University of Groningen (Study Approval Code: PSY-2122-S-0308). Half of these participants ($n = 52$) were randomly assigned to the first experiment, and the other half of these participants were assigned to the second experiment (see below).

Design and Procedure

This study used a two-condition, within-subjects design. Half of the participants were randomly assigned to a *pretesting* condition first, followed by a *prestudy* condition. For the other half of the participants, this order was reversed. In both conditions, participants studied 20 items (see the Materials section). These 20 items were divided into two 10-item subsets, one for each condition, through which participants cycled four times. Before each iteration, the order of the first five and the last five items was shuffled.

The trial sequence in both conditions is shown in Figure 1: Experiment 1. In the pretesting condition, participants were shown a country outline and were asked to type the associated country name on every repetition. If participants did not know the name of a country outline upon its first presentation, they were instructed to guess the correct answer. In the prestudy condition, for the first presentation of an item, participants were shown the correct answer, which they could simply reproduce in the response field. For all subsequent presentations, participants were asked to retrieve the answer. After typing the answer, participants received feedback ("Correct!" if the response was correct, "Incorrect, the correct answer was [correct answer]" if the response was incorrect, and "Too slow! The correct answer was [correct answer]" if the participants took more than 15 s to respond).

Materials

As one of the main aims of this study was to test the benefits of pretesting in applied settings, where learners are expected to have varying degrees of prior knowledge of study materials, we previously conducted a large-scale experiment to obtain prior knowledge norms for a set of materials, and we use those materials here with a comparable participant sample (Wilschut et al., 2023). In that work, we asked 287 participants (all first-year psychology students in the Netherlands, mainly with Dutch or German nationality) to name 114 country outlines. The country list contained the 100 most populous countries in the world and was supplemented with (a) countries in close proximity to the Netherlands (e.g., Luxembourg) and (b) countries in Europe with a fairly recognizable shape (e.g., Norway, Iceland). To obtain prior knowledge norms, the country outlines were presented to the participants three times, and participants were asked to name the countries (no feedback was given). For the present study, the 40 items with the highest average accuracy were chosen, resulting in an item set in which the most difficult item (Somalia) was correctly named by 11% of participants in the prior study, and the easiest item (Italy) was correctly named by 96% of participants. The average accuracy—a reliable estimation of the expected prior knowledge rate in the present study given the similarity of the participant samples—was 46%. The 40-item set was divided into two 20-item subsets, which were assigned to each experimental condition.[2] The country outline graphics used in this study were generated using R 3.4.1 (R Core Team, 2020) with packages *ggplot2* (Wickham, 2016) and *rnaturalearth* (South, 2017). The script for generating the images can be found online (see https://osf.io/uq8bw). For some countries, the images were manually edited after they were generated to fit the display (e.g., by removing a far outlying island).

Hardware and Software

The experiment was built with JavaScript and HTML5 using the *jsPsych* online experiment library (de Leeuw, 2015), and the experiment was conducted online.

Power Analysis

The sample size for this and the following studies was initially chosen based on the sample sizes used in other related studies (e.g., see Kornell & Vaughn, 2016). To confirm the power of this experiment and to establish that a sufficient number of participants was included in the study, we performed a power analysis. In the experiments reported here, we used (generalized) linear mixed-effects regression models to examine the differences between the prestudy and pretesting conditions. As conventional power analysis methods cannot be applied directly to such models (e.g., see Brysbaert & Stevens, 2018), we performed a series of simulation-based power analyses using the *simr* package (Green & MacLeod, 2016) in R 3.4.1 (R Core Team, 2020). Based on previous experiments, we expected an average learning accuracy of around 75% and mean response times (RTs) of around 3,000 ms in the adaptive-scheduling condition (see Experiments 2 and 4) and passive study condition (e.g., see Wilschut et al., 2024). We included an estimated effect of pretesting based on theoretical considerations: As we were only interested in detecting effects that would have a certain educational relevance, we wanted to have sufficient power to detect a difference between pretesting and prestudy of at least 8%–10% points at the first retrieval attempt following the initial trial. Finally, based on previous studies (e.g., see Wilschut et al., 2024), we expected a small increase in learning performance over item repetitions, as well as an interaction between the effect of pretesting versus prestudy (where the benefits of pretesting

[2] Due to a technical flaw in the experiment, the items were not randomly divided over the conditions. As a consequence, the average expected prior knowledge or difficulty of the items was not exactly the same in both experimental conditions. Therefore, we controlled for any possible effects of the difficulty of the items statistically (see Analyses section below).

would decrease over repetitions of the item). We simulated data for 50 repetitions of an experiment based on the above conditions with 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 participants and found an average power of 80% in detecting the effects of pretesting versus prestudy when at least 40–50 participants were included in the study. Therefore, the actual number of participants resulted in a sufficiently powered experiment. The full code and description of the power analysis are available online (see https://osf.io/3djyc).

Analyses

Data processing and statistical analyses were conducted in R 3.4.1 (R Core Team, 2020) with the mixed-effects modeling package *lme4* (Bates et al., 2012). The data were visualized using *ggplot2* (Wickham, 2016). We used (generalized) linear mixed-effects models to compare learning accuracy and RTs for each iteration of an item. For the first experiment, we fitted four mixed-effects models (see Table 1A–D). The first two models use accuracy as a dependent variable, and the latter two models fit RTs. For both accuracy and RTs, we fitted an "overall model," showing the overall effects of initial retrieval type, and a "split model," showing the effects of retrieval type, split by the success of the initial retrieval attempt. In more detail, the mixed-effects models reported in this study include as fixed effects (a) initial trial type (contrast coded, pretesting = 1; prestudy = 0); (b) iteration (0–2)[3]; (c) successful first iteration (contrast coded, successful first retrieval attempt = 1, unsuccessful first retrieval attempt = 0); and (d) the difficulty of the item, based on the prior knowledge norms (proportions between 0.11 for the most difficult item and 0.96 for the easiest item) for the materials (see Wilschut et al., 2023). Second-order interactions were included for all combinations of (a), (b), and (c). Accuracy (generalized binomial linear mixed-effects models) or log-transformed RTs (linear mixed-effects models) were dependent variables. Participant IDs were added as random intercepts to all models (Baayen et al., 2008). As the aim of this study was to examine the effects of the initial iteration on later retrieval performance, in all models, the first iteration (Repetition 0 in overall Models A and C or Retrieval Attempt 1 in split Models B and D) was not included in the analyses. All data related to this study, as well as analysis scripts, can be found in the additional online material (see https://osf.io/2wbqv).

Results

The main goal of the first experiment was to compare the effects of pretesting relative to prestudy on subsequent-repetition retrieval attempts. Figure 2 gives an overall summary of performance over repetitions in all four experiments. The results of Experiment 1 are shown in the first row (Figure 2A–2D). In the first two columns in Figure 2, accuracy is depicted on the *y*-axis. In the last two columns, RTs are shown on the *y*-axis. Figure 2A and 2C shows overall performance over item repetitions, and Figure 2B and 2D shows performance split by the success of the first retrieval attempt. Note that in the prestudy condition, the *second* presentation of an item (i.e., Repetition 1) corresponds to the first retrieval attempt, whereas in the pretesting condition, the *first* presentation of an item (i.e., Repetition 0) corresponds to the first retrieval attempt. There was no fourth retrieval attempt in the prestudy condition (the third repetition was the third retrieval attempt, whereas in the pretesting condition, the third repetition was the fourth retrieval attempt). Table 1 shows the results of 16 (generalized) linear mixed-effects models that correspond to the 16 subpanels shown in Figure 2: Table 1A corresponds to the data plotted in Figure 2A, Table 1B corresponds to Figure 2B, and so forth. The results of the first experiment can be summarized in five points.

First, Figure 2A shows, in line with earlier studies, that there was a large benefit of pretesting: Accuracy on the first repetition of an item was higher following the pretesting trial than following the prestudy trial. Table 1A shows that the mixed-effects model fit corroborates what is apparent in Figure 2A: The significant main effect of initial trial type indicates that accuracy after pretest was significantly higher than performance after prestudy. The same pattern of results was reflected in RTs, as shown in Figure 2C and Table 1C: Although attempting to retrieve an answer took more time compared to passively studying it (Repetition 0, not included in Table 1), participants were faster at later repetitions of the item if it was pretested (Repetitions 1 through 3).

Second, Figure 2A (Table 1A) and 2C (Table 1C) shows that accuracy increased and RTs decreased over repetitions following the first presentation of the item. This indicates that (unsurprisingly) performance improved over iterations (i.e., participants were successfully learning the materials).

Third, the interaction effects of initial trial type and iteration were significant, indicating that after the prestudy or pretesting trial, the size of the pretesting benefit *decreased* over iterations, both in terms of accuracy and in terms of RTs. Based on the coefficients from the mixed-effects regression models, we can calculate that after 1.6 iterations (0.82 [main effect of initial trial type]/0.51 [absolute interaction effect of initial trial type and iteration]), retrieval accuracy was expected to be the same in the prestudy and in the pretesting conditions. As iteration was referenced to 0 for Repetition 1, this means that the pretesting benefit disappeared after 2.6 repetitions (or 3.6 encounters including the initial prestudy/pretest trial). Figure 3 shows the net retrieval time benefit of using pretesting compared to using prestudy over iterations of an item. Circles represent average performance in the fixed-schedule experiment (Experiment 1), and triangles represent performance in the adaptive-scheduling experiment (Experiment 2; see below for details). The figure shows that for the fixed-schedule experiment, RTs were expected to converge after 4 (0.20/0.05, see Table 1C) iterations (five repetitions or six encounters including the initial pretest/prestudy trial). The total time cost of pretesting was nullified by the added average time gain on later repetitions. In short, we found that the pretesting benefit (higher accuracy after pretesting than after prestudy) was largest upon the first repetition following the initial presentation of the item and progressively got smaller until disappearing after approximately three repetitions of an item.

Fourth, the interaction effect between the success of the first retrieval attempt and the initial trial type was *not* significant with respect to accuracy (see Table 1B). Visually, Figure 2B appears to show a larger benefit of pretesting following an unsuccessful retrieval attempt than after a successful retrieval attempt. However, when

[3] The term "iteration" was used to create a general term for *item repetition* in the overall models and for *retrieval attempt* in the split models. Note that, for the split models, this first retrieval attempt corresponded to the first presentation of the item in the pretesting condition (the first trial was a retrieval trial) but to the second presentation of an item in the prestudy condition (here, the second trial was the first retrieval trial).

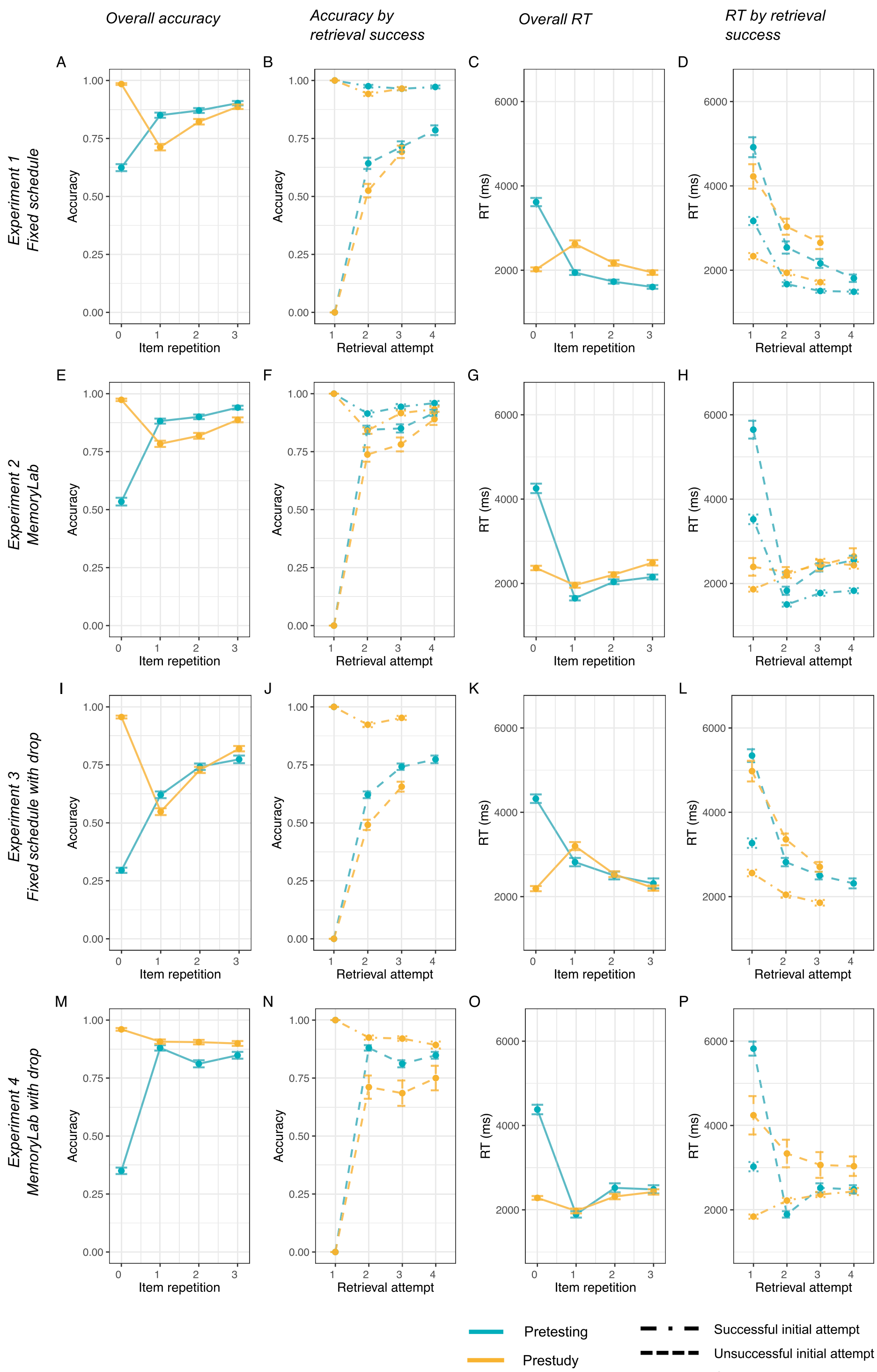
Table 1
Summary of Mixed-Effects Model Results for Accuracy and Response Times in All Experiments

Experiment	Factor	Overall accuracy	Accuracy by retrieval success	Overall RT (log[ms])	RT by retrieval success (log[ms])
Model 1 (Fixed, no drop)		A	B	C	D
	Intercept	1.16***	0.96***	7.73***	7.75***
	Initial trial type (pretest = 1, prestudy = 0)	0.82***	0.42*	−0.20***	−0.18***
	Iteration	0.81***	1.02***	−0.13***	−0.12*
	Initial Trial Type × Iteration	−0.51***	−0.51*	0.05**	−0.01
	Successful First Iteration × Initial Trial Type		0.22		0.15**
Model 2 (MemoryLab, no drop)		E	F	G	H
	Intercept	1.22***	1.10***	7.55***	7.63***
	Initial trial type	0.63***	0.65***	−0.07***	−0.15***
	Iteration	0.36***	0.37***	0.05***	0.05***
	Initial Trial Type × Iteration	−0.14*	−0.15	0.00	0.03
	Successful First Iteration × Initial Trial Type		−0.22		−0.02
Model 3 (Fixed, drop)		I	J	K	L
	Intercept	−0.26*	−0.96***	7.96***	8.00***
	Initial trial type	0.31***	1.12***	−0.27***	−0.33***
	Iteration	0.78***	0.76***	−0.16***	−0.20**
	Initial Trial Type × Iteration	−0.28***	−0.27	0.07**	0.12
Model 4 (MemoryLab, drop)		M	N	O	P
	Intercept	2.01***	0.46	7.59***	7.85***
	Initial trial type	−0.50***	0.94***	−0.17***	−0.50***
	Iteration	−0.04	0.21*	0.06***	−0.02
	Initial Trial Type × Iteration	0.12*	−0.10	0.03**	0.10***

Note. Models A–P refer to 16 separate mixed-effects regression models and correspond to the 16 subpanels in Figure 2. Factors can be interpreted as follows. Initial trial type: Overall effect of pretesting versus prestudy—is there a pretesting benefit? Iteration: Overall effect of number of iterations—how does performance change over repetitions (overall models) or retrieval attempts (split models)? Initial Trial Type × Iteration: Change of the effect of initial trial type over iterations—how does the pretesting benefit change over iterations? Successful First Iteration × Initial Study Method: Effect of initial retrieval success on the effect of initial trial type—is there a difference in pretesting benefit after successful compared to unsuccessful pretesting? Note that Repetition 0 (overall Models A, C, E, G, I, K, M, and O) or Retrieval Attempt 1 (split Models B, D, F, H, J, L, N, and P) were not included in the analyses. The iteration variable is referenced to 0 for the first repetition (overall models) or the second retrieval attempt (split models). For the split models in Experiments 3 and 4, models were run on data for trials following successful initial retrieval attempts only. Only a selected number of effects of interest are shown. RT = response time.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Figure 2

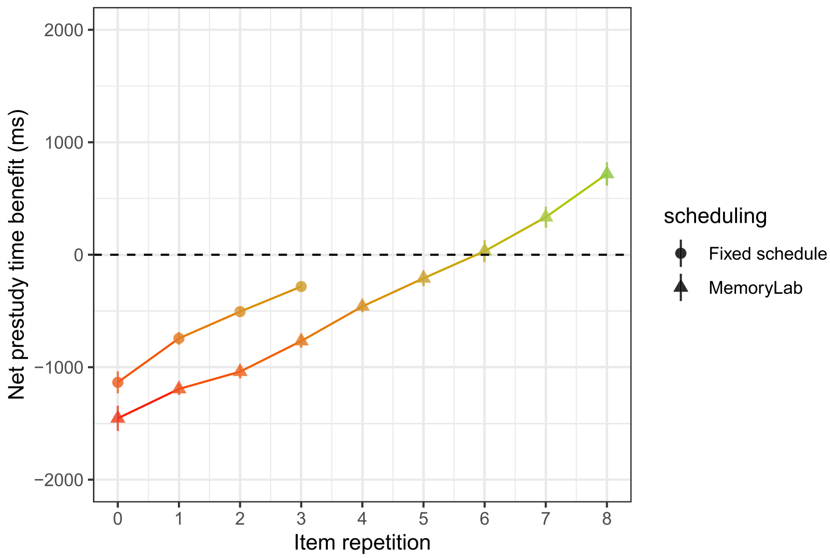
Retrieval Performance Over Iterations



Note. The first column (A, E, I, M) shows overall mean accuracy in the pretesting (blue) and prestudy (yellow) conditions. The second column (B, F, J, N) shows mean accuracy separated by the success of the first retrieval attempt. The third column shows overall mean response times (C, G, J, K, O) and the fourth column (D, H, L, P) shows response times separated by the success of the first retrieval attempt. Error bars show (±1) standard error of the mean. The four rows correspond to the four experiments. RT = response time. See the online article for the color version of this figure.

Figure 3

Cumulative Net Difference in Total Retrieval Time After Pretesting Compared to Prestudy Over Item Repetitions



Note. Negative values represent a greater net retrieval time in the pretesting condition than in the prestudy condition, and positive values represent more time invested in the prestudy compared to the pretesting condition. Circles correspond to the fixed item scheduling (Experiment 1), and triangles correspond to MemoryLab adaptive item scheduling (Experiment 2). Vertical lines show (±1) standard error of the mean. See the online article for the color version of this figure.

accuracy scores are compared using a generalized linear model, the same difference on the logit scale translates to changes of different magnitudes depending on where they happen on the percentage scale. Correspondingly, the logistic mixed-effects models demonstrate that there was no difference in the size of the pretesting benefit after successful versus unsuccessful retrieval attempts.

Finally, we found that if a pretest is correct, accuracy remained high at around 97% (see Figure 2B, top blue line). This is important because it suggests that after a successful initial retrieval attempt, it can be safely assumed that the learner does not need to keep practicing the item and thus that the item could be removed from the set, leaving more time to study other items (this idea is tested in Experiments 3 and 4).

In summary, the results of the first experiment show that pretesting boosted retrieval performance in a realistic, multiple-repetition retrieval practice session. However, the benefit was largest upon the first repetition of the item and progressively diminished over repetitions. We found no effect of the success of the first retrieval attempt on the size of the prestudy benefit.

The previous analyses show a clear benefit of pretesting over prestudy on performance at the retrieval practice trials directly following the initial trial. Figure 4 shows the benefits of pretesting relative to prestudy by learner. The average proportion of correct answers in the pretesting condition is shown on the y-axis, and the proportion of correct answers in the prestudy condition is shown on the x-axis. The subpanels show sequential repetitions of the items. The left subpanel shows the first item presentation. Here, performance was much higher in the prestudy condition than in the pretesting condition, because in the prestudy condition the correct answer was shown on the screen. For the first and second item repetition, there was a clear pretesting benefit: Most learners were above the diagonal that indicates equal performance in both conditions. On the third repetition, accuracy increased for most learners, and the difference between pretesting and prestudy got smaller. Overall, the figure shows that the pretesting benefits shown in Figure 2 (first row) were not driven by a specific subset of learners but found in the participant pool as a whole.

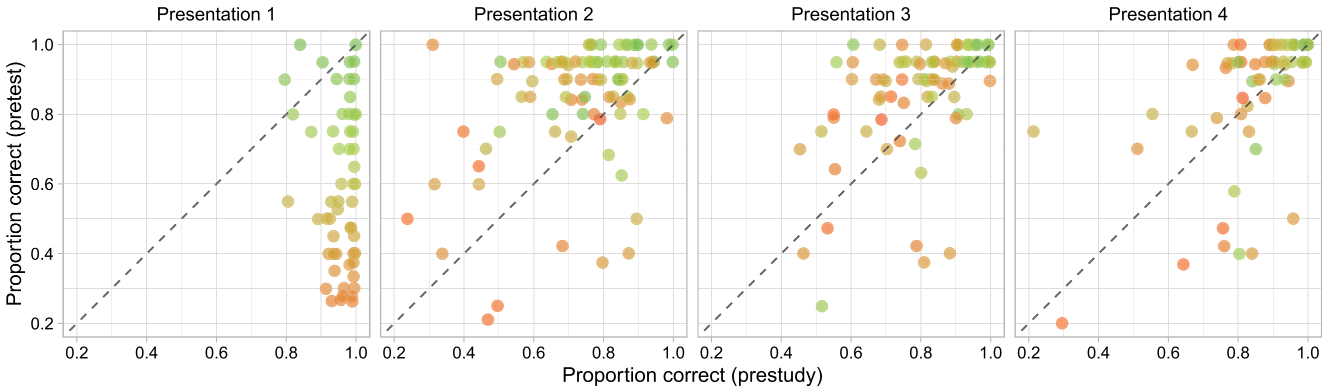
Experiment 2: Pretesting Benefits in Adaptive Retrieval Practice

Method

The second experiment used the same materials, hardware and software, and analysis methods as Experiment 1. The study was performed using the second half of the participants that were invited for Experiments 1 and 2 (participants completed either Experiment 1 or 2). As we did not have any upfront expectations for how learning behavior in the fixed-schedule experiment would differ from the learning performance in the adaptive-scheduling conditions, we used the same power analysis for both the fixed-schedule and adaptive-scheduling conditions. The second experiment differed from the first in the way in which the items were scheduled. In Experiment 1, the items were scheduled using a fixed item repetition schedule. In Experiment 2, we used the *MemoryLab* adaptive-scheduling algorithm that is used in a wide range of real-world educational contexts (Sense et al., 2016; van der Velde, Sense, Spijkers, et al., 2021; van Rijn et al., 2009; see https://MemoryLab.nl/en). MemoryLab is based on the Adaptive Control of Thought–

Figure 4

Pretesting Benefits Over Iterations



Note. The *x*-axis shows the proportion of correct responses in the prestudy condition, and the *y*-axis shows the proportion of correct responses in the pretesting condition. Dots are individual learners. Each subpanel represents a subsequent presentation of the items. Dots falling above the diagonal line represent better average performance in the pretesting condition than in the prestudy condition. The color of the item represents the average accuracy in the pretesting trial. This figure shows the results of Experiment 1. See the online article for the color version of this figure.

Rational architecture's model of human declarative memory (Anderson et al., 1998; Pavlik & Anderson, 2008), and it aims to identify the best item repetition schedules for individual learners by a model-tracing process based on RTs and accuracy scores. By having the learner rehearse each item just before it is estimated to be forgotten, the algorithm balances the beneficial effects of active retrieval during learning and spacing learning over time (Cepeda et al., 2008; Karpicke & Bauernschmidt, 2011; Kornell, 2009; Nakata, 2017). In practice, this means that easier items are repeated later and less often than more difficult items and that faster learners encounter more items than slower learners. The beneficial effects of using adaptive item repetition schedules have been demonstrated both in controlled lab studies (Mettler et al., 2011; Sense et al., 2016; van der Velde, Sense, Borst, & van Rijn, 2021; Zhou et al., 2021) and in real-world classroom situations (Sense et al., 2021; van der Velde, Sense, Spijkers, et al., 2021; van Rijn et al., 2009): Learning facts with the MemoryLab system resulted in around 8%–10% better recall of studied materials compared to learning with less adaptive, accuracy-based algorithms in which RTs were not taken into account (van Rijn et al., 2009). Of particular importance to the present study, the way in which the MemoryLab algorithm presents items resembles a frequently used realistic learning context in which items are rehearsed based on the needs of the learner and do not follow a fixed-repetition structure (see Mettler et al., 2016).

In Experiment 1 (fixed-schedule retrieval practice), participants completed an 80-trial sequence. Because the MemoryLab system adaptively adjusts the number of items and the number of times each item should be repeated, in Experiment 2, we did not set the number of trials to a fixed amount. Instead, the study duration was set to 9 min (for comparison, the average duration in the fixed-schedule learning condition was 8.1 min).

Results

The main aim of the second experiment was to examine whether the effects found in Experiment 1 generalized to a more realistic learning context, where items do not follow a fixed-repetition schedule but are dynamically scheduled based on their difficulty and based on the individual's learning pace. In such a context, the number of repetitions per item will vary between items and learners, which means that any pretesting benefit is also likely to vary more than in Experiment 1. In addition, since the MemoryLab algorithm presents items when their activation reaches a threshold, early repetitions of a fact tend to follow each other quickly, while later repetitions tend to be spaced further apart. The accuracy and RTs over item repetitions for the pretesting and prestudy conditions are shown in Figure 2E–2H. Again, the results can be summarized in five points.

First, as in the first experiment, pretesting resulted in higher average accuracy and lower average RTs compared to prestudy; see Table 1E and Table 1G, respectively. Second, we found that the effects of iteration are less pronounced compared to the effects found in the first experiment: Accuracy increased slightly over iterations, but RTs got slightly longer over iterations. This is in line with our expectations: The scheduling model is deliberately flattening these curves by repeating items on a schedule that ideally leads to stable repetition-to-repetition performance. Third, as in Experiment 1, there was a significant interaction between initial trial type and iteration in the overall accuracy model, indicating that the size of the pretesting benefit decreased over iterations. The estimated point at which accuracy in both conditions converges was after 4.5 repetitions (0.63/0.14). The interaction effect between initial trial type and iteration was not significant in the overall RT model, indicating the RT benefit after pretesting did not change over repetitions. Fourth, as in the first experiment, the interaction between the success of the first retrieval attempt and the effect of the initial trial type was *not* significant, indicating that there was no difference in pretesting benefit after successful versus unsuccessful first retrieval attempts. Finally, we found that after a correct initial retrieval attempt, accuracy dropped slightly at the second retrieval attempt, but then remained stable and high (see Figure 2F, top blue line), indicating that also in adaptive-scheduling sessions, items that are correctly retrieved may not need to be kept in the practice pool.

In summary, the results of the second experiment, in which a personalized item scheduling system was used, align with the results of the first experiment. In both experiments, we found a significant benefit of pretesting on accuracy for the following retrieval attempts that gets smaller over iterations. Both experiments did not show a significant difference in the size of the pretesting benefit between successful and unsuccessful attempts. Additionally, both experiments showed that for items that were correctly recalled in the pretesting trial, accuracy remained high at later repetitions. The results differed in the effects of iteration: In the adaptive-scheduling condition, performance remained more constant over repetitions, whereas performance improved over iterations in the fixed scheduling condition.

Interim Discussion

The aim of the first two experiments was to examine the benefits of pretesting relative to prestudy prior to a multiple-repetition retrieval practice session. The results from both experiments provide evidence for the beneficial effects of pretesting in applied settings: Accuracy was higher and RTs were lower at retrieval attempts immediately following pretesting than for retrieval attempts following prestudy. The accuracy of the pretesting trial was not an important predictor of the size of the pretesting benefit, indicating that even when the learner did not know the answer to an item, attempting to retrieve it facilitated learning.

Despite the fact that we found a clear benefit of pretesting on retrieval performance in the first two or three retrieval attempts immediately following the initial trial, the benefit of pretesting diminished over repetitions: The expected point at which retrieval performance in both conditions (both in terms of retrieval accuracy and in terms of RT) converged is after approximately three to four repetitions. Since realistic learning usually involves multiple repetitions for each item in a learning session (e.g., see Mettler et al., 2016; Nakata, 2011; Sense et al., 2021), this finding calls into question the practical usefulness of pretesting when followed by retrieval practice. We also found that for items that were correctly recalled at a pretesting trial, accuracy remained high at later repetitions, suggesting that participants already had relatively stable memory representations for these items at the pretesting trial. In Experiments 3 and 4, we aimed to use pretesting trials to identify these memory-stable items and subsequently explore whether personalizing the schedule by dropping these items, and replacing them with new items, would result in enhanced overall learning outcomes.

Experiment 3: Prior Knowledge–Based Personalization in Fixed-Schedule Retrieval Practice

Method

Experiment 3 used the same hardware and software and analysis methods as Experiment 1. The study was carried out using a new participant pool, consisting of 106 first-year psychology students (71 female, 35 male) at the University of Groningen who participated in either Experiment 3 or 4. The mean age of these participants was 21.5 ($SD = 4.3$) years. As in the first two experiments, most of the participants were Dutch or German. Half of these participants ($n = 53$) were randomly assigned to complete Experiment 3; the other half was assigned to Experiment 4. We used the same procedure to perform a power analysis as in Experiments 1 and 2, but we used slightly different values for the expected accuracy and RTs. As we used a prior knowledge–based item drop in this experiment (see below), we expected that the accuracy on the remaining items would decrease, as the easier or known items would be filtered out of the set. Therefore, we performed a new power analysis with a base accuracy of 65% instead of 75% and a base RT of 3,500 ms instead of 3,000 ms in the passive study condition. All other values remained the same. We found an average power of 80% in detecting the effects of pretesting versus prestudy, with at least 50 participants (for more details see https://osf.io/3djyc). Therefore, the actual sample size resulted in a sufficiently powered experiment. As in Experiments 1 and 2, participants gave informed consent before starting, and the study was approved by the ethical committee of the department of psychology at the University of Groningen (Study Approval Code: PSY-2122-S-0308B).

In Experiment 3, items were scheduled based on the same fixed-repetition structure as used in Experiment 1, with one exception: Items were removed from the learning set if they were retrieved correctly at initial retrieval in the pretesting condition. For each dropped item, a new item would be introduced (which, in turn, could also be dropped if named correctly at the first presentation). In the prestudy condition, no items were dropped, and no new items were introduced. To ensure a sufficiently large item pool with the new item drop manipulation, Experiment 3 used a larger (N items = 74) and more difficult country outline subset than Experiments 1 and 2 (see https://osf.io/uq8bw). Unlike in Experiments 1 and 2, a final test in which participants were tested on all encountered items followed immediately after the learning session.

Results

Learning Performance

Figure 2I and 2J shows accuracy during learning in Experiment 3. Note that the prestudy condition in Experiment 3 was similar to the prestudy condition in Experiment 1. In the pretesting condition, items were dropped after correct pretesting and replaced with a new item. Therefore, Figure 2J and 2L does not show accuracy and RTs for successful initial attempt trials in the pretesting condition. The results of this experiment can be summarized in four points.

First, overall accuracy (Figure 2I; Table 1I) was slightly higher after pretesting than after prestudy, indicating that even when known items were replaced with new items in the pretesting condition, there was a pretesting benefit for the items that were initially answered incorrectly. Similarly, overall RTs (Figure 2K; Table 1K) were lower after a failed pretesting than after prestudy. Second, there was an effect of iteration on both overall accuracy and overall RTs, showing that accuracy increased and RTs decreased over repetitions. Third, in the overall models, the interaction effect of iteration and initial trial type was significant. This indicates that, as in Experiment 1, the size of the pretesting benefit, both in terms of accuracy and in terms of RTs, diminished over repetitions. Finally, Figure 2J (Table 1J) and 2L (Table 1L) shows that, as in the first two experiments, we found no significant differences in the benefits of pretesting as a function of initial trial success, which shows that even when *known* items are replaced by new items, attempting to retrieve *unknown* items was beneficial.

Test Performance

As the main aim of this experiment was to explore the benefits of prior knowledge–based personalization in terms of learning outcomes, we asked participants to complete a cued recall test after the learning session in which they were asked to name all countries they encountered in the learning session. First, Figure 5A shows the test performance on items that were dropped during the learning session for fixed item scheduling (Experiment 3) and MemoryLab item scheduling (Experiment 4). In the fixed scheduling experiment, in 94.2% of all cases, dropped items were correctly recalled on the test. In the MemoryLab scheduling experiment, dropped items were correctly recalled on the test in 97.7% of all cases. Overall, as test accuracy on dropped items was very high, these results show that dropping the items after an initial correct response was justified. Figure 5B shows that average performance on items that were studied was much lower than average performance on dropped items. Finally, Figure 5C highlights two of the most frequently made errors on dropped items on the test. For example, the most frequently made error for "Niger" was "Peru," and vice versa, indicating that participants often interchanged answers for the two country outlines. Along similar lines, "Venezuela" was the most common error for the country outline of "Estonia," and vice versa. Overall, test performance on dropped items was high, and the few mistakes that were made generally seemed to have been a consequence of the visual similarity of pairs of country outlines.

Figure 6A shows test performance for individual participants as a function of their prior knowledge. Here, prior knowledge is defined as the number of correct pretesting attempts in the pretesting condition. Table 2A shows the linear mixed-effects model results associated with Figure 6A. At zero prior knowledge, the main effect of initial trial type was not significant, indicating that at the low end of the prior knowledge scale, there was no advantage, but also no *dis*advantage, of pretesting in combination with item replacement relative to prestudy. In other words, even when all items are unknown to the learner, attempting to retrieve them does not impede learning in terms of the overall number of items correctly retrieved at the test. We found a significant main effect of prior knowledge, demonstrating that the number of items that were recalled on the test was higher if prior knowledge was high. Finally, there was a significant interaction effect of initial trial type and prior knowledge: For each item a learner knew prior to the learning session, there was a 0.61 item increase in overall learning benefit associated with using pretesting *plus* item replacement relative to using the prestudy condition. This means that at median prior knowledge (nine items), we found on average ($9 \times 0.61 - 2.84 =$) 2.65 more correct responses on the test.

In summary, we found that dropping items after correct pretesting was justified, as follow-up test performance on dropped items was very high. Dropped items for which incorrect responses *were* given on the test were often visually similar to other items in the set. We also show that using prior knowledge–based personalization (i.e., replacing items that are retrieved correctly in a pretesting trial with new items) can increase overall learning efficiency. Importantly, also for participants with low prior knowledge, pretesting-based personalization of the item set did not lead to a lower learning efficiency.

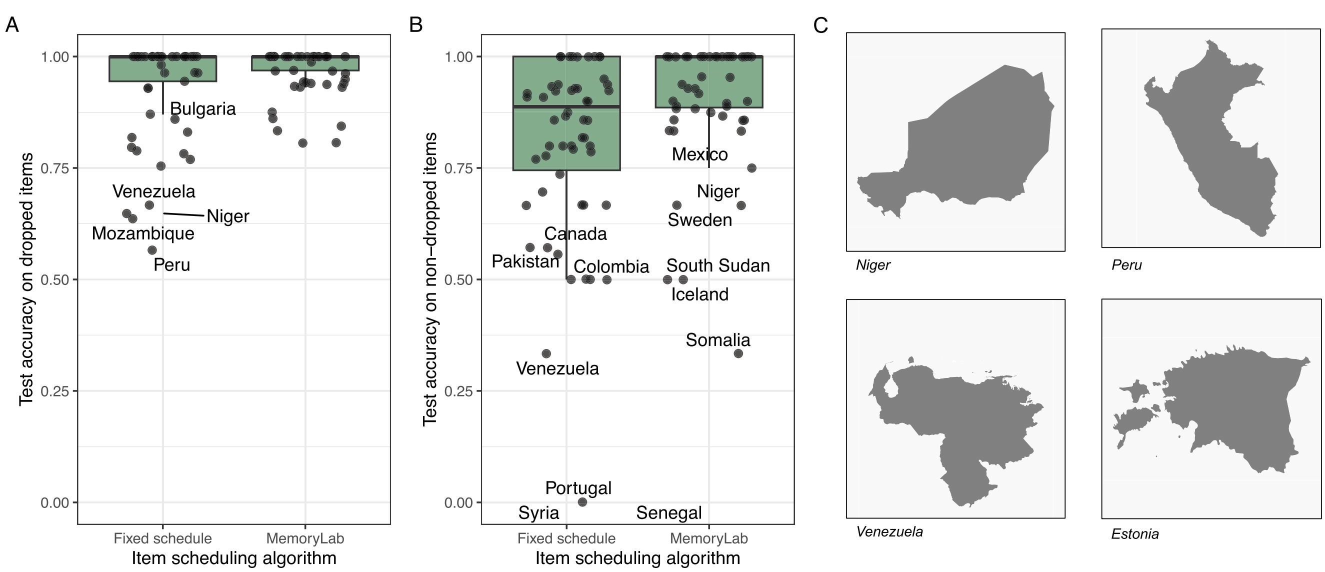
Experiment 4: Prior Knowledge–Based Personalization in Adaptive Retrieval Practice

Method

Experiment 4 used the same study materials, hardware and software, and analysis methods as Experiment 2. The study was

Figure 5

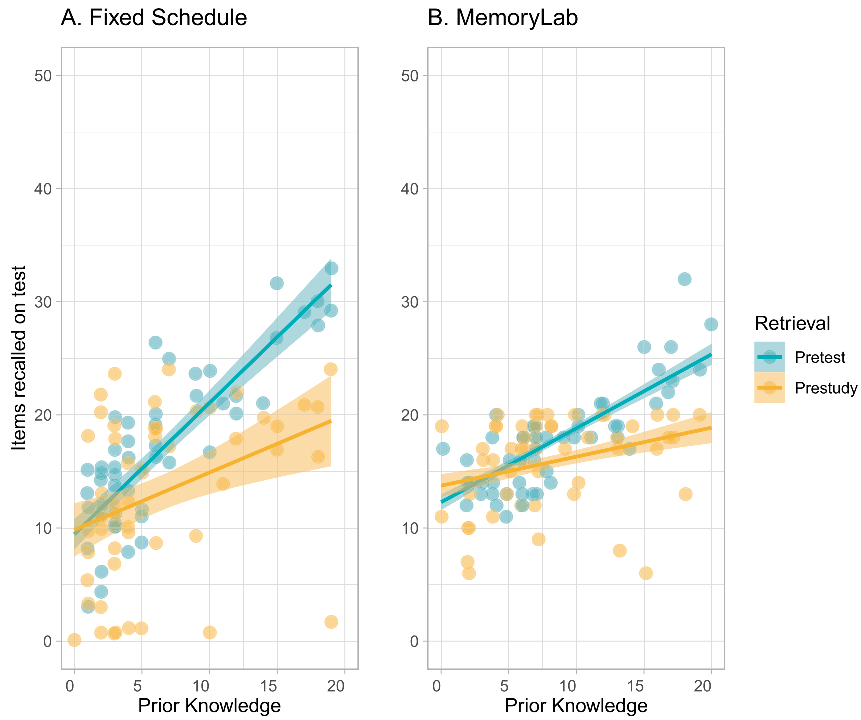
Test Performance on Items Dropped During Learning



Note. (A) The average accuracy on test for dropped items in the fixed-schedule experiment (Experiment 3) and the MemoryLab experiment (Experiment 4). (B) The average accuracy on the test for nondropped, studied items. (C) Two examples of frequent mistakes, where participants mistakenly swapped responses for "Niger" and "Peru," and "Venezuela" and "Estonia." See the online article for the color version of this figure.

Figure 6

Number of Country Outlines Named Correctly on Test



Note. (A) Fixed-schedule learning (Experiment 3). (B) MemoryLab learning (Experiment 4) as a function of the amount of prior knowledge. Dots show individual participants. See the online article for the color version of this figure.

performed using the second half of the participants that were invited for Experiments 3 and 4, and we used the same power analysis as in Experiment 3. Experiment 4 used the same adaptive-scheduling system as Experiment 2. In addition, it used the same “drop” manipulation as Experiment 3: Country outlines that were correctly named at initial retrieval in the pretesting condition were removed from the set and replaced with a new item. The learning session had a set duration of 12 min in both the pretesting and prestudy conditions. As in Experiment 3, after the learning session, a test followed in which participants were tested on all encountered items.

Results

Learning Performance

The aim of the fourth and final experiment was to test the benefits of prior knowledge–based personalization, as explored in Experiment 3, in an adaptive learning session. Figure 2M and 2N shows accuracy during learning in Experiment 4, and Figure 2O and 2P shows RTs. We found a significant disadvantage in overall learning accuracy after pretesting compared to prestudy; see Table 1M. In other words, we found that overall learning performance in the pretesting condition *from which known items were dropped* was slightly lower compared to performance in the prestudy condition *including known items*. It is good to point out that prestudy accuracy in this experiment was relatively high compared to the other experiments, which might be part of the reason why we do not find a benefit of prestudy over pretesting in this experiment. We did find an overall benefit of pretesting in terms of RTs; see Table 1O. As in Experiment 3, there were benefits of pretesting after unsuccessful retrieval attempts, both in terms of accuracy (Table 1N) and RTs (Table 1P). The effects of iteration were—as in Experiment 2—less pronounced in Experiment 4 than in Experiment 3, indicating that the adaptive learning algorithm successfully kept performance constant over iterations.

Test Performance

Figure 6B (Table 2B) shows test performance after a retrieval practice session with pretesting and after prestudy. In both conditions, the number of to-be-studied items was determined by the MemoryLab system based on RTs and accuracy scores. As in Experiment 3, we found no significant difference in the number of studied items in the pretesting and prestudy conditions at zero prior knowledge. In addition, we again found an effect of prior knowledge: The more items were known prior to the experiment, the higher the final score on the test. Finally, the interaction between initial trial type and prior knowledge is significant, indicating that the more prior knowledge a learner has, the bigger the benefit of using pretesting in combination with item replacement is. At median prior knowledge (nine items), there was a 6.83 item benefit of using pretesting and prior knowledge–based item scheduling. In summary, these results demonstrate that using a realistic learning approach, prior knowledge–based personalization using pretesting trials boosts learning efficiency, especially

Table 2
Summary of Mixed-Effects Model Results for the Number of Recalled Items on Test

Experiment	Factor	*N* correct on test
Model 1 (Fixed, drop)		A
	Intercept	13.73***
	A. Initial trial type (pretest = 1, prestudy = 0)	−2.84
	B. Prior knowledge	0.25**
	C. Initial Trial Type × Prior Knowledge	0.61***
Model 2 (MemoryLab, drop)		B
	Intercept	10.46***
	A. Initial trial type	−2.08
	B. Prior knowledge	0.37**
	C. Initial Trial Type × Prior Knowledge	0.99***

Note. Models A and B refer to two separate linear mixed-effects regression models and correspond to the two subpanels (A and B, respectively) in Figure 4. Factors can be interpreted as follows. Initial trial type: Overall effect of pretesting versus prestudy: Is there a pretesting benefit at zero prior knowledge? Prior knowledge: Overall effect of prior knowledge: Do learners with more prior knowledge score higher on the test? Initial Trial Type × Prior Knowledge: Effect of prior knowledge on the effect of initial trial type: Is there a larger pretesting benefit for learners with more prior knowledge? Only a selected number of effects of interest are shown in this table.
** $p < .01$. *** $p < .001$.

when learners had a moderate to high level of prior knowledge on the subject.

Transparency and Openness

In the Method sections, as well as in the additional online material, we report how we determined our sample size, all data exclusions, all manipulations, and all measures for the present four experiments. All data, analysis code, and research materials are available online (see https://osf.io/2wbqv). Details on the data analyses are provided in the Analyses section. This study's design and its analysis were not preregistered.

Discussion

The main goal of this study was to explore the extent to which pretesting compared to prestudy prior to a multiple-repetition retrieval practice session can improve overall learning outcomes. To this end, we conducted a set of experiments with materials constructed to ensure varying degrees of prior knowledge on the study materials, items were repeated multiple times after initial presentation, and learners received continuous feedback throughout the session. We examined (a) the beneficial effects of pretesting on subsequent retrieval practice trials and (b) overall learning efficiency if pretesting trials are used to identify and remove known items from the learning set (Experiments 3 and 4). The results of this study can be summarized in three main points, the first two concerning (a) and the third regarding (b).

Pretesting Benefits Multiple Repetition Retrieval Practice

First, we found strong evidence for benefits of pretesting on subsequent retrieval practice performance: Both successful and unsuccessful initial retrieval attempts resulted in improved retrieval accuracy on retrieval attempts directly following the initial trial. One deviation to this pattern was the overall effect of pretesting on accuracy in Experiment 4, where overall accuracy was lower after pretesting than after prestudy. However, the latter comparison directly contrasts a condition in which all items are included (prestudy) to a condition in which all known items are dropped (pretesting), making the overall learning difficulty much higher. Regardless of the success of the initial retrieval attempt and regardless of the item scheduling, RTs for retrieval attempts following pretesting were lower than RTs for retrieval attempts following prestudy, providing additional support for the idea that pretesting results in stronger memory representations (Anderson & Schooler, 1991; van Rijn et al., 2009). In line with earlier research, we found that the accuracy on the pretest did not meaningfully influence the size of the pretesting benefit, suggesting that the act of trying to retrieve the answer to a question, and not successful retrieval as such, enhanced later retrieval performance. Overall, these findings show that pretesting has a beneficial effect on later retrieval practice performance, both when a controlled item repetition schedule is used and when a more realistic adaptive item repetition schedule is used.

Second, we found that these pretesting benefits on retrieval practice performance (accuracy and RTs) got progressively smaller over item repetitions. Where many studies demonstrated the effects of pretesting on single study-test trials (e.g., see Arnold & McDermott, 2013; Izawa, 1970), we were interested in the effects of an *initial* pretesting versus prestudy trial on retrieval performance over multiple retrieval attempts. Importantly, the conditions differed only in the first trial: In both conditions, the initial trial was followed by three active retrieval practice trials and feedback. We found that the benefit of pretesting persisted, but progressively got smaller, over trials. In general, the above results are consistent with the mechanisms underlying pretesting effects as proposed by Pan and Carpenter (2023). First, it is plausible that presenting the prestudy trials increased the learners' general curiosity during the task, resulting in better encoding of the cue-answer pair when the feedback on the pretesting trial was presented. Subsequently, this better encoding of the cue-answer pair can have enhanced retrieval performance at the retrieval practice trials following the initial trial. It is not surprising that the benefits of initial pretesting get smaller over repetitions, as the factors that induce the learners' curiosity during the pretesting trial are also present during the following retrieval practice trials. In fact, it is interesting to note that even after a retrieval practice trial (i.e., a trial that is exactly the same as the initial pretesting trial), the benefits of *starting* with a testing trial are still present, suggesting that sparking the learners' curiosity at the beginning of a session by presenting a question is more effective than quizzing the learner after the answer has already been presented passively. In addition to affecting the psychological state of the learner, the formation of stronger memories of the cue during the presentation of the pretesting trial might have positively affected retrieval during the following retrieval practice trials. Again, it is reasonable to expect that the benefit of pretesting over prestudy decreases repetitions following the initial pretesting trial, because such cue-specific memories are also formed during the retrieval practice trials that nullify the effect of the initial trial.

We find these effects both when a fixed item repetition schedule is used and when a more realistic, performance-based adaptive item scheduling system is used. We included the performance-based scheduling approach to mimic realistic learning scenarios, where a learner chooses to allocate practice time to items depending on their difficulty (e.g., see Mettler et al., 2016). In the adaptive-scheduling condition, we find smaller effects of iterations: Performance remains relatively stable throughout the session. This is expected: The adaptive item scheduling system increases the spacing of items that are already known (making the task more difficult) and repeats more difficult items quickly (making the task easier). This results in a key difference compared to Experiment 1: It helps prevent motivational issues and boredom as a consequence of items that are too easy or frustration when items are too difficult (e.g., see Minear et al., 2018; Moeyaert et al., 2016).

Overall, the results of Experiments 1 and 2 show that after about four repetitions, performance in both initial trial type conditions was expected to converge, calling into question the practical usefulness of pretesting in multiple-repetition learning sessions. In order to examine whether pretesting trials could be used to enhance overall learning efficiency in applied settings, we tested the possibility of exploiting pretesting trials to facilitate prior knowledge–based personalization in Experiments 3 and 4.

Pretesting Can Be Used to Identify a Learner's Prior Knowledge

Third, we showed that pretesting could be used to identify prior knowledge and personalize the item repetition schedule based on this information. After a correct initial retrieval attempt, accuracy remained high on later repetitions (see Figure 2B and 2F). Furthermore, Figure 5A shows that dropped items were correctly recalled on the test in 94%–98% of all cases. In other words, if a learner knows the answer to a question prior to learning, our data suggest that it could be removed from the set of to-be-learned items, potentially leading to benefits in overall learning efficiency. Further inspection of the items that were answered incorrectly on the test after correct initial pretesting suggests that errors on dropped items were frequently a result of a participant swapping two visually similar country outlines (e.g., see Figure 5B). Further studies should examine the possibility of presenting a few additional retrieval questions for to-be-dropped items during the learning session—especially if the item set contains items that are similar—to ensure that the participant knows the item. Overall, the results of Experiments 3 and 4 confirmed the idea that pretesting trials can be used to identify prior knowledge and successfully optimize overall learning efficiency for individual learners based on this information. Both when item scheduling was fixed and when an adaptive-scheduling system was used, dropping items after successful pretesting resulted in a higher number of total items learned. We showed that the additive benefits of employing such prior knowledge–based personalization are most pronounced for participants with high levels of prior knowledge.

Although we found that prior knowledge–based personalization seemed to be most beneficial for learners with a certain amount of prior knowledge, it is important to stress that pretesting did not impair learning if prior knowledge was low. The Appendix figure shows an additional analysis in which we show accuracy and RTs over repetitions for all participants (first row), participants with low prior knowledge (second row), and participants with middle–high prior knowledge (third row). In line with earlier results by Buchin and Mulligan (2023), who found that prior knowledge did not influence the benefits of pretesting on subsequent retrieval practice, we found the same general pattern of results in low prior knowledge participants and in middle–high prior knowledge participants. In all prior knowledge groups, performance on item repetitions following the pretesting trial was better than performance on item repetitions following a prestudy trial. In other words, *even if a learner did not know the materials*, the cost associated with attempting to retrieve the answer to each item was nullified by enhanced performance on the item repetitions immediately following the retrieval attempt, leading to overall similar learning efficiency in both initial study conditions. A noticeable deviation in the response pattern for low prior knowledge participants relative to middle–high prior knowledge participants, however, is the approximately 5% point drop in accuracy after a successful initial retrieval attempt (see Appendix Figure F). In the low prior knowledge group, a successful pretesting trial did not necessarily indicate that an item is stably stored in memory (e.g., it could also reflect a correct guess on the first attempt or a lack of engagement with the task during the second retrieval attempt). These observations can be construed as another case for reintroducing to-be-dropped items a few more times after pretesting, before permanently dropping it and replacing it with a new item. One open question concerns the way in which the prior knowledge–based personalization affects the learners' affective or metacognitive states (Hamari et al., 2016; Kennedy et al., 2014; Shernoff et al., 2003). More specifically, it is possible that for learners with low prior knowledge, having to produce answers to retrieval questions might be demotivating, and future studies should examine ways to counter this issue.

In short, regardless of prior knowledge, pretesting directly enhances retrieval performance on the next couple of item repetitions following the initial retrieval attempt. Furthermore, we show that pretesting trials can be used to identify prior knowledge and subsequently personalize item repetition schedules based on the results. Such prior knowledge–based personalization is especially effective when a learner is partly familiar with the study materials before learning but does not negatively impact learners with little or no prior knowledge.

Limitations

The results of this study should be interpreted with some caution. First, as the study was conducted online, there may be concerns about the validity of the data and the engagement of the participants. However, we programmatically monitored whether participants were engaged with and focused on the task (e.g., they could not click away without notice), RTs did not show unusual values, and performance increased over iterations, suggesting focused learning. Second, the exact type of materials used is likely to play an important role in examining the benefits of pretesting. In this study, we chose country outlines because participants could be expected to have varying degrees of prior knowledge on the topic. As a consequence, they will—at least to some extent—have felt capable of successfully retrieving the names of some of the country outlines in the pretesting condition, which may be a necessary condition for initial retrieval attempts to be beneficial. Future studies are needed to validate our results using different types of stimuli. A potential confound to consider is whether possible motivational decreases over the course of the session may have resulted in diminishing benefits of pretesting over repetitions. In particular, it is important to note that in the current experimental setting, the learner's motivation to complete the task differs from educational practice, where behavior may be guided by more intrinsic goal orientations or self-control (e.g., Duckworth & Gross, 2014). However, we found that overall accuracy increases and RTs decrease as the session progresses, indicating continued engagement and learning throughout the task (see Figure 2). Additionally, the task itself is relatively short (on average, only 8.1 min for Experiment 1), a duration typically insufficient to induce fatigue or a notable decline in motivation (e.g., see Ackerman & Kanfer, 2009; Trejo et al., 2005). Despite these points, future studies should address potential confounds of motivation by explicitly measuring it.

Implications

Our results have evident implications for the further development of digital tools that aid the memorization of facts through retrieval practice. We show that pretesting enhanced such retrieval practice under realistic digital learning conditions, though with the caveat that performance benefits disappeared with more than a few repetitions. Furthermore, we show that pretesting was an efficient strategy when pretesting trials were used to identify and remove known items from the learning set. In educational contexts, learners may already be familiar with some of the learning materials prior to the learning session. Importantly, we did not find an overall disadvantage of pretesting, even if participants had a low amount of prior knowledge of the study materials. The study was not conducted in a classroom setting, so the findings may not fully generalize to direct educational contexts. However, it was carried out online with first-year undergraduates who could complete the task at their convenience, reflecting real-world digital learning habits. The materials were relevant, and participants were motivated by the opportunity to earn course credits. In this way, the study design mirrors digital learning applications used in the real world.

We think that our results can be particularly useful in the context of computerized adaptive learning systems as they show that pretesting trials can be used as a reliable tool to identify prior knowledge. Current adaptive learning systems typically face the problem of trying to determine what prior knowledge the learner already has and require an initial study and test phase in order to determine appropriate feedback or select useful practice problems. This is called the cold-start problem: Having learners study and complete tests takes time, which makes it difficult for the system to properly adjust to the situation early in the learning session (e.g., see Pliakos et al., 2019; van der Velde, Sense, Borst, & van Rijn, 2021). The results of this study suggest that replacing initial study and test trials with active retrieval attempts may be an efficient way of countering an aspect of this issue: If a learner cannot correctly answer an item, the pretest retrieval attempt will promote learning, and time costs associated with the pretest retrieval attempt will be compensated by faster RTs in later repetitions of the item. If the learner does know the answer, the item may be (temporarily) removed from the set of to-be-studied items, leaving the learner more time to practice unfamiliar materials.

In conclusion, we show that pretesting, relative to prestudy, can improve retrieval performance in a multiple-repetition retrieval practice session, where learners receive continuous feedback and have varying degrees of prior knowledge of the topic. We found these effects both in a fixed-schedule learning session and in an adaptive learning session in which the item presentation was adjusted to the needs of the individual learner and both when the pretests were answered successfully and when they were answered incorrectly. After approximately four repetitions, the benefit of pretesting diminished. However, if pretesting trials were used to select and replace known items for individual learners, it proved to be an efficient overall study strategy. These results may guide the development of computerized, adaptive learning applications that aim to enhance memorization in educational practice.

References

Ackerman, P. L., & Kanfer, R. (2009). Test length and cognitive fatigue: An empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology: Applied*, *15*(2), 163–181. https://doi.org/10.1037/a0015719

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*(3), 659–701. https://doi.org/10.3102/0034654316689306

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

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, *38*(4), 341–380. https://doi.org/10.1006/jmla.1997.2553

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*(6), 396–408. https://doi.org/10.1111/j.1467-9280.1991.tb00174.x

Arnold, K. M., & McDermott, K. B. (2013). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 940–945. https://doi.org/10.1037/a0029199

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., Krivitsky, P. N., Tanaka, E., Jagan, M., & Boylan, R. D. (2012). *lme4: Linear mixed-effects models using "Eigen" and S4* [Computer software]. R Foundation for Statistical Computing. https://cran.r-project.org/web/packages/lme4/index.html

Brod, G. (2021). Toward an understanding of when prior knowledge helps or hinders learning. *npj Science of Learning*, *6*(1), Article 24. https://doi.org/10.1038/s41539-021-00103-w

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), Article 9. https://doi.org/10.5334/joc.10

Buchin, Z. L., & Mulligan, N. W. (2023). Retrieval-based learning and prior knowledge. *Journal of Educational Psychology*, *115*(1), 22–35. https://doi.org/10.1037/edu0000773

Carpenter, S. K. (2023). Encouraging students to use retrieval practice: A review of emerging research from five types of interventions. *Educational Psychology Review*, *35*(4), Article 96. https://doi.org/10.1007/s10648-023-09811-8

Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology*, *1*(9), 496–511. https://doi.org/10.1038/s44159-022-00089-1

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

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*, *19*(11), 1095–1102. https://doi.org/10.1111/j.1467-9280.2008.02209.x

de Leeuw, J. R. (2015). jspsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Dochy, F., de Rijdt, C., & Dyck, W. (2002). Cognitive prerequisites and learning: How far have we progressed since bloom? Implications for educational practice and teaching. *Active Learning in Higher Education*, *3*(3), 265–284. https://doi.org/10.1177/1469787402003003006

Duckworth, A., & Gross, J. J. (2014). Self-control and grit: Related but separable determinants of success. *Current Directions in Psychological Science*, *23*(5), 319–325. https://doi.org/10.1177/0963721414541462

Geller, J., Toftness, A. R., Armstrong, P. I., Carpenter, S. K., Manz, C. L., Coffman, C. R., & Lamm, M. H. (2018). Study strategies and beliefs about learning as a function of academic achievement and achievement goals. *Memory*, *26*(5), 683–690. https://doi.org/10.1080/09658211.2017.1397175

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

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

Hailikari, T., Nevgi, A., & Lindblom-Ylänne, S. (2007). Exploring alternative ways of assessing prior knowledge, its components and their relation to student achievement: A mathematics based case study. *Studies in Educational Evaluation*, *33*(3–4), 320–337. https://doi.org/10.1016/j.stueduc.2007.07.007

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, *54*, 170–179. https://doi.org/10.1016/j.chb.2015.07.045

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology*, *83*(2, Pt. 1), 340–344. https://doi.org/10.1037/h0028541

Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 576–597). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.028

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, *93*(3), 579–588. https://doi.org/10.1037/0022-0663.93.3.579

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review*, *27*(2), 317–326. https://doi.org/10.1007/s10648-015-9309-3

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1250–1257. https://doi.org/10.1037/a0023436

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772–775. https://doi.org/10.1126/science.1199327

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*(5865), 966–968. https://doi.org/10.1126/science.1152408

Kennedy, P., Miele, D. B., & Metcalfe, J. (2014). The cognitive antecedents and motivational consequences of the feeling of being in the zone. *Consciousness and Cognition*, *30*, 48–61. https://doi.org/10.1016/j.concog.2014.07.007

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, *23*(9), 1297–1317. https://doi.org/10.1002/acp.1537

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

Latimier, A., Peyre, H., & Ramus, F. (2021). A meta-analytic review of the benefit of spacing out retrieval practice episodes on retention. *Educational Psychology Review*, *33*(3), 959–987. https://doi.org/10.1007/s10648-020-09572-8

Latimier, A., Riegert, A., Peyre, H., Ly, S. T., Casati, R., & Ramus, F. (2019). Does pre-testing promote better retention than post-testing? *npj Science of Learning*, *4*(1), Article 15. https://doi.org/10.1038/s41539-019-0053-1

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

Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology*, *68*, 465–489. https://doi.org/10.1146/annurev-psych-010416-044022

Mettler, E., Massey, C., & Kellman, P. (2011). *Improving adaptive learning technology through the use of response times*. Proceedings of the Annual Meeting of the Cognitive Science Society. https://escholarship.org/uc/item/2xs4n8wz

Mettler, E., Massey, C. M., & Kellman, P. J. (2016). A comparison of adaptive and fixed schedules of practice. *Journal of Experimental Psychology: General*, *145*(7), 897–917. https://doi.org/10.1037/xge0000170

Minear, M., Coane, J. H., Boland, S. C., Cooney, L. H., & Albat, M. (2018). The benefits of retrieval practice depend on item difficulty and intelligence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(9), 1474–1486. https://doi.org/10.1037/xlm0000486

Moeyaert, M., Wauters, K., Desmet, P., & Van den Noortgate, W. (2016). When easy becomes boring and difficult becomes frustrating: Disentangling the effects of item difficulty level and person proficiency on learning and motivation. *Systems*, *4*(1), Article 14. https://doi.org/10.3390/systems4010014

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning*, *24*(1), 17–38. https://doi.org/10.1080/09588221.2010.520675

Nakata, T. (2017). Does repeated practice make perfect? The effects of within-session repeated retrieval on second language vocabulary learning.

Studies in Second Language Acquisition, *39*(4), 653–679. https://doi.org/10.1017/S0272263116000280

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

Pan, S. C., & Sana, F. (2021). Pretesting versus posttesting: Comparing the pedagogical benefits of errorful generation and retrieval practice. *Journal of Experimental Psychology: Applied*, *27*(2), 237–257. https://doi.org/10.1037/xap0000345

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied*, *14*(2), 101–117. https://doi.org/10.1037/1076-898X.14.2.101

Pliakos, K., Joo, S.-H., Park, J. Y., Cornillie, F., Vens, C., & Van den Noortgate, W. (2019). Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems. *Computers & Education*, *137*, 91–103. https://doi.org/10.1016/j.compedu.2019.04.009

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

R Core Team. (2020). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, *140*(3), 283–302. https://doi.org/10.1037/a0023956

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

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

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Sense, F., Behrens, F., Meijer, R. R., & van Rijn, H. (2016). An individual's rate of forgetting is stable over time but differs across materials. *Topics in Cognitive Science*, *8*(1), 305–321. https://doi.org/10.1111/tops.12183

Sense, F., van der Velde, M., & van Rijn, H. (2021). Predicting university students' exam performance using a model-based adaptive fact-learning system. *Journal of Learning Analytics*, *8*(3), 155–169. https://doi.org/10.18608/jla.2021.6590

Shernoff, D. J., Csikszentmihalyi, M., Shneider, B., & Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly*, *18*(2), 158–176. https://doi.org/10.1521/scpq.18.2.158.21860

Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2022). Domain-specific prior knowledge and learning: A meta-analysis. *Educational Psychologist*, *57*(1), 31–54. https://doi.org/10.1080/00461520.2021.1939700

Soderstrom, N. C., & Bjork, E. L. (2023). Pretesting enhances learning in the classroom. *Educational Psychology Review*, *35*(3), Article 88. https://doi.org/10.1007/s10648-023-09805-6

South, A. (2017). *rnaturalearth: World map data from natural earth* (R package Version 0.1.0) [Computer software]. https://doi.org/10.32614/CRAN.package.rnaturalearth

Thompson, R. A., & Zamboanga, B. L. (2003). Prior knowledge and its relevance to student achievement in introduction to psychology. *Teaching of Psychology*, *30*(2), 96–101. https://doi.org/10.1207/S15328023TOP3002_02

Trejo, L. J., Kochavi, R., Kubitz, K., Montgomery, L. D., Rosipal, R., & Matthews, B. (2005). *Measures and models for predicting cognitive fatigue*. Proceedings of SPIE-Biomonitoring for Physiological and Cognitive Performance During Military Operations. https://doi.org/10.1117/12.604286

van der Velde, M., Sense, F., Borst, J., & van Rijn, H. (2021). Alleviating the cold start problem in adaptive learning using data-driven difficulty estimates. *Computational Brain & Behavior*, *4*(2), 231–249. https://doi.org/10.1007/s42113-021-00101-6

van der Velde, M., Sense, F., Spijkers, R., Meeter, M., & van Rijn, H. (2021). Lockdown learning: Changes in online foreign-language study activity and performance of Dutch secondary school students during the COVID-19 pandemic. *Frontiers in Education*, *6*, Article 712987. https://doi.org/10.3389/feduc.2021.712987

van Rijn, H., van Maanen, L., & van Woudenberg, M. (2009). *Passing the test: Improving learning gains by balancing spacing and testing effects* [Conference session]. Proceedings of the 9th International Conference of Cognitive Modeling.

Vaughn, K. E., & Rawson, K. A. (2012). When is guessing incorrectly better than studying for enhancing memory? *Psychonomic Bulletin & Review*, *19*(5), 899–905. https://doi.org/10.3758/s13423-012-0276-0

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer.

Wilschut, T., Sense, F., & van Rijn, H. (2024). Speaking to remember: Model-based adaptive vocabulary learning using automatic speech recognition. *Computer Speech & Language*, *84*, Article 101578. https://doi.org/10.1016/j.csl.2023.101578

Wilschut, T., van der Velde, M., Sense, F., & van Rijn, H. (2023). Prior knowledge norms for naming country outlines: An open stimulus set. *Journal of Cognition*, *6*(1), Article 14. https://doi.org/10.5334/joc.260

Witherby, A. E., & Carpenter, S. K. (2022). The rich-get-richer effect: Prior knowledge predicts new learning of domain-relevant information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(4), 483–498. https://doi.org/10.1037/xlm0000996

Yan, V. X., Yu, Y., Garcia, M. A., & Bjork, R. A. (2014). Why does guessing incorrectly enhance, rather than impair, retention? *Memory & Cognition*, *42*(8), 1373–1383. https://doi.org/10.3758/s13421-014-0454-6

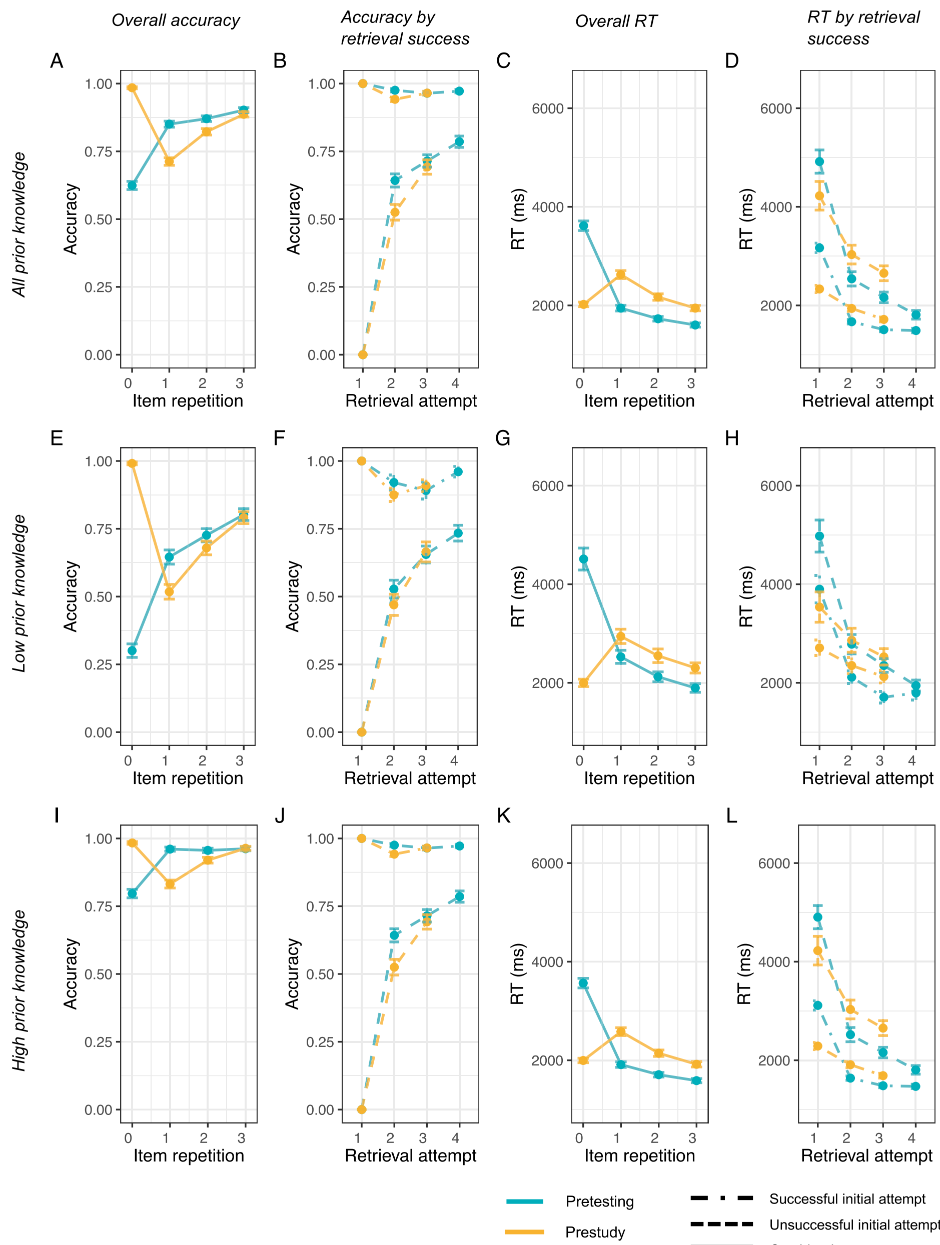
Zambrano, J., Kirschner, F., Sweller, J., & Kirschner, P. A. (2019). Effects of prior knowledge on collaborative and individual learning. *Learning and Instruction*, *63*, Article 101214. https://doi.org/10.1016/j.learninstruc.2019.05.011

Zhou, P., Sense, F., van Rijn, H., & Stocco, A. (2021). Reflections of idiographic long-term memory characteristics in resting-state neuroimaging data. *Cognition*, *212*, Article 104660. https://doi.org/10.1016/j.cognition.2021.104660

(*Appendix follows*)

Appendix

Retrieval Performance Over Iterations for Experiment 1, Separated for Three Prior Knowledge Groups



Note. The first column (A, E, I) shows overall accuracy in the pretesting (blue) and prestudy (brown) conditions. The second column (B, F, J) shows accuracy separated by the success of the first retrieval attempt. The third column (C, G, K) shows overall response times, and the fourth column (D, H, L) shows response times separated by the success of the first retrieval attempt. The first row shows learners in all prior knowledge groups, and the second row summarizes response data of the learners with the lowest prior knowledge (lowest quartile). The final row shows the participants with the highest prior knowledge (highest quartile). RT = response time. See the online article for the color version of this figure.

Received April 15, 2024
Revision received March 20, 2025
Accepted April 22, 2025 ■