



OPEN ACCESS

EDITED BY

David Clewett,
University of California, Los Angeles,
United States

REVIEWED BY

Emily Cowan,
Temple University, United States
James Robert Houston,
Middle Tennessee State University,
United States

*CORRESPONDENCE

Dagmar Zeithamova
✉ dasa@uoregon.edu

RECEIVED 14 July 2023

ACCEPTED 26 September 2023

PUBLISHED 12 October 2023

CITATION

Ashby SR and Zeithamova D (2023) A combination of restudy and retrieval practice maximizes retention of briefly encountered facts. *Front. Cognit.* 2:1258955. doi: 10.3389/fcogn.2023.1258955

COPYRIGHT

© 2023 Ashby and Zeithamova. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A combination of restudy and retrieval practice maximizes retention of briefly encountered facts

Stefania R. Ashby^{1,2} and Dagmar Zeithamova^{3*}

¹Department of Psychology, Brigham Young University, Provo, UT, United States, ²Neuroscience Center, Brigham Young University, Provo, UT, United States, ³Department of Psychology, University of Oregon, Eugene, OR, United States

Introduction: Is retrieval practice always superior to restudy? In a classic study by Roediger and Karpicke, long-term retention of information contained in prose passages was found to be best when opportunities to restudy were replaced with opportunities to self-test. We were interested whether this striking benefit for repeated testing at the expense of any restudy replicates when study opportunities are brief, akin to a single mention of a fact in an academic lecture. We were also interested in whether restudying after a test would provide any additional benefits compared to restudying before test.

Method: In the current study, participants encountered academically relevant facts a total of three times; each time either studied (S) or self-tested (T). During study, participants predicted how likely they were to remember each fact in the future. During self-test, participants performed covert cued recall and self-reported their recall success. Final test followed immediately or after a delay (Experiment 1: 2 days, Experiment 2: 7 days).

Results: Contrary to prior work, long-term memory was superior for facts the were restudied in addition to self-tested ($SST > STT = SSS$). We further investigated whether restudy after a test (STS) provides additional benefits compared to restudy before test (SST). Restudying after a retrieval attempt provided an additional benefit compared to restudying before a retrieval attempt on an immediate test, but this benefit did not carry over a delay. Finally, exploratory analyses indicated that restudy after test improved the accuracy of participants' subjective predictions of encoding success.

Discussion: Together, our results qualify prior work on the benefits of repeated testing, indicating that balancing testing with repetition may allow for more information to be learned and retained. These findings offer new insights into the conditions that promote encoding and long-term retention, provide new constraints for existing cognitive theories of testing effects, and have practical implications for education.

KEYWORDS

retrieval practice, testing effects, long-term memory, forgetting, metamemory

Introduction

Factors that help us encode information into long-term memory and reduce forgetting have long been of major interest in memory research. For example, seminal work by Ebbinghaus (1913) examined how restudying material affected long-term memory retention and forgetting over time. He showed that the rate of forgetting is especially rapid early after learning, but repeated study reduces forgetting. Many have examined this repetition effect in other memory paradigms and confirmed that increasing study

repetitions results in more accurate recall (Mayer, 1983; Bromage and Mayer, 1986; Webb, 2007) and recognition (Davis et al., 1972; Donaldson, 1981) memory performance.

In addition to repetition, retrieval practice has also been shown to boost memory performance (Gates, 1917; Spitzer, 1939; Tulving, 1967; Hogan and Kintsch, 1971). A common way to induce retrieval practice is by testing the to-be-remembered material after its initial study, with the term *retrieval practice effect* or *testing effect* referring to the increased memory retention after retrieval compared to studying alone. Retrieval practice benefits memory, even when it comes at the expense of repetition. For example, Roediger and Karpicke (2006a) had participants either study prose passages twice (SS) or study once and practice retrieval once via free recall (ST) without corrective feedback. After a delay, participants were given a final free-recall test to assess long-term retention. After intervals of 2-days and 1-week, participants who practiced retrieval had better memory retention than participants who studied twice as much without an opportunity for retrieval practice (ST > SS). This phenomenon has been replicated numerous times, demonstrating that retrieval of material with and without feedback strengthens subsequent memory (Roediger and Butler, 2011; Rowland, 2014; for review and meta-analyses see Adesope et al., 2017).

When there are multiple opportunities for study or test, repeated testing seems to provide additional benefits. In the same study by Roediger and Karpicke (2006a), participants were given prose passages and instructed to read and study for the duration of a 5 min period. Additional 5 min periods were filled with either more study time (S) or with test time (T) where participants tried to retrieve as much of the passage as possible without an opportunity to re-read it or receive feedback on their recall. After a 5 min delay, participants who only had the initial study period followed by multiple test periods (STTT) recalled the least amount of detail, with more study time being associated with better performance (STTT < SSST < SSSS). In contrast, when retrieval was tested after a week-long delay, the pattern of results reversed, with repeated testing being associated with better performance than repeated study (STTT > SSST > SSSS). Furthermore, it appeared that repeated testing also minimized forgetting—comparable performance was seen in the STTT condition tested after 5 min as when tested after a week, while performance was lower after a week in all other conditions (Roediger and Karpicke, 2006a). Thus, although a limited amount of information was encoded in the single study period, as evidenced by lower STTT performance in the 5 min delay condition, almost all that was encoded was resistant to forgetting over a period of a week, due to repeated testing.

But should all study repetition be replaced by testing to maximize long-term memory? While past work has shown that replacing repetition with testing better prevents forgetting of the initially learned items, there are limitations. Given that only a fraction of information will be remembered after a single encounter, one could conclude that including some repetition may provide greater memory retention overall as much may be missed with such little initial exposure. Indeed, it is important to note that even in the single study conditions discussed above, participants typically had multiple opportunities to study each piece of information. For example, in Roediger and Karpicke (2006a), the study period was fixed at 5 min rather than one read-through of the passage. Thus, participants had an opportunity to re-read the text passage several

times within that period. In Karpicke and Roediger (2007), another study that found a benefit of repeated testing over repetition, each word-list study cycle consisted of five passes through the list, so participants in the single study condition (STTT) actually studied each wordlist five times (and 15 times in the SSST condition). Thus, the *combination* of repetition and retrieval practice in these conditions may be critical for long-term retention, rather than repeated testing alone. The first goal of the current study was to test to what degree replacing repetition with repeated retrieval practice affects learning and retention under more extreme conditions where material is briefly encountered, akin to a mention of a fact during an academic lecture. Because limited information can be remembered after a single exposure, we hypothesized that balancing testing with repetition may be more beneficial than multiple test opportunities.

To test this hypothesis, we conceptually followed the seminal study by Roediger and Karpicke (2006a), testing immediate and delayed memory after varying repetition and retrieval attempts while keeping the overall time with the material constant. Here, participants encountered academically relevant stimuli (facts from a range of disciplines) three times each, either complete (study, S) or with a missing word that they covertly attempted to retrieve (self-test, T). The degree of study and self-test opportunities varied across stimuli in a within-subject design, such that after the initial study, some of the facts were restudied twice (SSS condition), some were restudied once and self-tested once (SST condition) and some were not restudied but instead were self-tested twice (STT condition). As feedback would provide an extra study opportunity (e.g., making an SST condition effectively an SST(S) condition) and make interpretation unclear (Greving and Richter, 2018), potentially conflating retrieval practice effects with test-potentiated learning effects (Arnold and McDermott, 2013), no feedback was provided (Nungester and Duchastel, 1982; Roediger and Karpicke, 2006a; Verkoeijen et al., 2012). To assess the effects of retrieval practice and repetition on learning and long-term retention, half of the participants completed the final test on the same day, while the other half completed the final test after a delay of 2 days (Experiment 1) or 1 week (Experiment 2).

In addition to improving retention, testing has also been reported to boost the efficiency of subsequent opportunities to restudy material (Izawa, 1966, 1970; Karpicke and Roediger, 2007; Kornell et al., 2009; Hays et al., 2013). This *test-potentiated learning effect*, or *indirect testing effect*, is present even when a previous retrieval attempt is unsuccessful and when no corrective feedback on the retrieval is provided (Arnold and McDermott, 2013). For this reason, Arnold and McDermott (2013) argue that studies showing greater retrieval practice effects when including feedback may be instead demonstrating the powerful combined effects of both retrieval practice effects and test-potentiated learning effects. To explicitly assess test-potentiated learning, several studies have intentionally incorporated conditions that include restudy after test (Karpicke and Roediger, 2007; van Gog and Kester, 2012; Dirkx et al., 2014). However, Arnold and McDermott (2013) argue that test-potentiated learning in some studies may be difficult to disentangle from retrieval benefits *per se*. For example, Dirkx et al. (2014) found better delayed performance in STST compared to SSSS, which could be due to retrieval practice or enhanced restudying of material after testing. Karpicke and Roediger (2007) held the number of study and retrieval opportunities equal between

conditions and still found a benefit for restudy after test (STST > SSTT). However, the benefit of the STST condition could be both due to more efficient restudy after test (enhanced second study) and because more material was strengthened by a second retrieval attempt after restudying (additional higher benefit of the second test).

Importantly, even without the conflation with the final retrieval attempt, one may expect that the order of restudy and test may affect long-term memory. Metacognitive awareness of encoding success, measured as the correspondence between subjective prospective memory judgments (i.e., judgments of learning) at study and objective retrieval success at test, tends to improve with the inclusion of a test (Koriat et al., 2002; Metcalfe and Finn, 2008b). Therefore, retrieval attempts, even when unsuccessful, may provide greater insight into what is known and what is still unknown, resulting in an opportunity to make subsequent study more effective (Kornell et al., 2009; Richland et al., 2009; Kornell, 2014). This suggests there may be additional benefits for an STS over an SST study protocol.

The order of study and test opportunities should also matter when one considers the bifurcation model of testing effects, proposed by Kornell et al. (2011). In the bifurcation model, restudy leads to a small increase in memory strength for all material, resulting in a shift of the whole distribution of memory strengths. In contrast, testing leads to a higher increase of memory strength, but only for items that are already above retrieval threshold, leading to a “bifurcated” (split) distribution of memory strength as items currently below retrieval threshold remain unstrengthened. Because the retrieved memoranda end up far above threshold (much further than they would be after re-study), they will tend to remain above threshold longer than restudied ones. Within this framework, both SST and STS training would be expected to lead to bifurcated distributions, but with a greater proportion of items in the high memory strength (bifurcated) part of the distribution in the SST condition. This would potentially lead to greater resistance to delay-related performance decline in the SST than STS condition.

Thus, the second objective of the current study was to examine the potential cost or benefit of restudying material after retrieval practice. To achieve this objective, we compared immediate and delayed performance in the SST condition with an STS condition, where a restudy opportunity followed a retrieval opportunity. Because retrieval attempts can make subsequent study more effective (Kornell et al., 2009; Richland et al., 2009; Kornell, 2014), we hypothesized that restudying after testing (STS) may boost learning and long-term retention to a greater degree than restudying before testing (SST). Finally, we hypothesized that restudy after test may also be accompanied by better metamemory awareness of one’s encoding success, as measured by the correspondence between trial-by-trial prospective memory judgments (reported during study) and final test performance (see Lovelace, 1984; Nelson and Narens, 1994; Kao et al., 2005).

Method

The two experiments described below had identical methods except for the length of delay in the Delay condition (Experiment 1 = 2 days; Experiment 2 = 7 days). Because the pattern of results was

the same across experiments, we report both experiments jointly for brevity and clarity, but the same conclusions would be reached by considering each experiment separately.

Participants

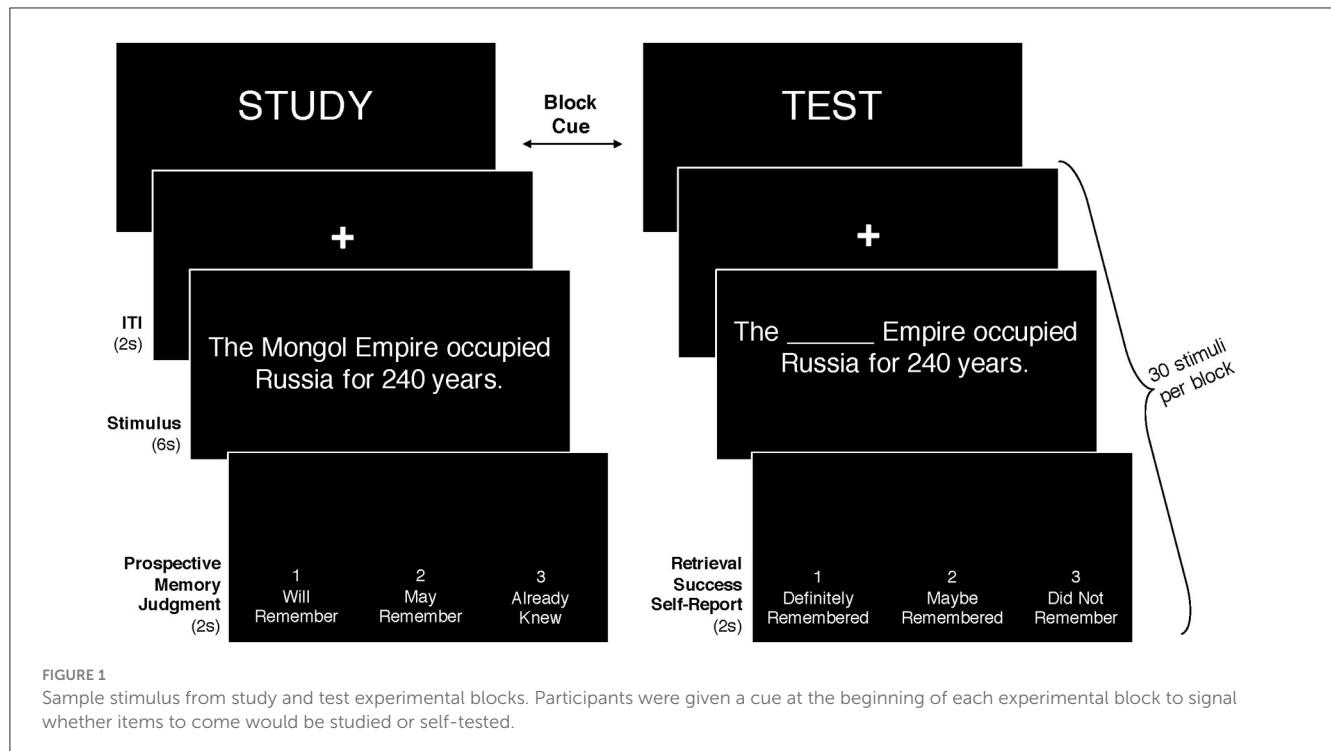
Participants were recruited from the University of Oregon community via the university SONA research system. Data were collected from 66 participants in Experiment 1 and 72 participants in Experiment 2. All participants provided written informed consent and received US \$25 compensation for their participation across two experimental sessions. All experimental procedures were approved by Research Compliance Services at the University of Oregon. Participants self-reported that they were either native or fluent English speakers and participants recruited for Experiment 2 were verified to have not participated in Experiment 1. One participant was excluded from each experiment for poor performance, indicated by overall accuracy three standard-deviations below the group mean. This left 65 participants from Experiment 1 ($M_{\text{age}} = 21.9$, $SD_{\text{age}} = 3.6$, age range: 18–33, 41 females, 24 males) and 71 participants from Experiment 2 ($M_{\text{age}} = 20.1$, $SD_{\text{age}} = 2.3$, age range: 18–32, 54 females, 17 males) for analyses.

Stimuli

Factual statements were used as stimuli to examine the retrieval-practice effects in an academically relevant context. Each statement had two formats: a study format and a test format. Statements in the study format consisted of complete, factual statements aggregated by the research team from a range of disciplines: art, astronomy, biology, chemistry, geography, health, world history, literature, music, mythology, neuroscience, perception, physical science, religion, sports, theater, and world travel. Statements in the test format were the same as those used in study except that one keyword from the original study statement was replaced with a “_____” (blank). See Figure 1 for an example study and test trial. Test statements were used to induce retrieval practice during the training portion of the experiment and used in the final test phase described below. For a given fact, the same blank was used anytime the fact appeared in the test form to avoid learning during test. For example, if we used “The Mongol empire occupied _____ for 240 years” during retrieval practice but “The _____ empire occupied Russia for 240 years” in a later test, the earlier test trial could end up serving as an additional study trial.

Procedure

All participants enrolled in a two-session experiment with a 2-day (Experiment 1) or 7-day (Experiment 2) delay between sessions. Participants first underwent training during Session 1, learning a series of facts presented on a computer screen in MATLAB using The Psychophysics Toolbox (Brainard, 1997).



From a larger bank of 202 factual statements, 150 statements were randomly selected for each participant and randomly assigned to one of four training conditions or one “never-studied” condition (30 statements per condition). Statements assigned to the never-studied condition were not presented during training; instead, their inclusion in the final test served two purposes. First, it allowed us to estimate individuals’ prior knowledge, providing a baseline. Second, as the immediate group would be taking their final test at the end of an hour-long experimental session and the delay group would be taking their final test on a new day without any prior testing, comparing performance between groups for the never-studied items would allow us to control for any group differences unrelated to learning, such as fatigue.

The training consisted of 12 blocks (Figure 2). Each block contained either 30 study or 30 self-test trials from one condition, with all trials separated by a 2 s interval. During *study* trials (Figure 1, left), participants were shown a factual statement and asked to study and memorize the sentence to the best of their ability (6 s). Participants were then presented with two response options on the screen (2 s), where they were asked to make a prospective memory judgment (e.g., “will remember” or “may remember”) indicating their confidence in recalling the statement later. Early pilot testing indicated that the observational nature of the task resulted in fatigue and participant disengagement, thus prospective memory judgements were primarily included to facilitate engagement with the task. Subsequently these ratings were utilized in an exploratory analysis of participants awareness of their encoding success (e.g., Kao et al., 2005), testing the degree to which the trial-by-trial prospective memory judgments (reported during study) corresponded to the actual retrieval success at final test. A third response option (“already knew”) was included during the initial study phase so that participants could indicate

if they already knew the presented fact prior to participating in the experiment. This option was included to allow us to estimate prior knowledge and focus analyses on facts learned during the experiment. During *test* trials (Figure 1, right), participants were shown the same statements (6 s), but with one of the key terms missing and replaced by a “_____” (blank). While some work suggests that testing effects are larger using overt retrieval practice techniques (Jonsson et al., 2014; Tauber et al., 2018), several studies also demonstrate that overt and covert retrieval methods provide similar benefits to long-term retention (Putnam and Roediger, 2013; Smith et al., 2013; Sundqvist et al., 2017; Tauber et al., 2018). We thus opted for a covert approach and asked participants to self-test and mentally fill-in-the-blank by retrieving the correct keyword from memory, without vocalizing or writing down their response. Participants then indicated how successful they were in recalling the missing keyword (2 s) by choosing from three response options: definitely remembered, maybe remembered, did not remember. Subsequent analyses relating these self-reported success responses and the actual final test performance confirmed that participants were compliant and accurate in their self-reports.

In all training conditions, each statement was encountered three times in a study or self-test format. Trials were blocked by condition and repetition (Figure 2). The first presentation of a statement was always in a study format. Training conditions differed in the format of subsequent encounters of the statements assigned to that condition. After the first study block of each condition was completed, training continued with a second block from each condition (in a study or test format) and then third block of each condition (in a study or test format). In the SSS condition, all blocks were in study format; in the SST condition, the last study block was replaced by a self-test block; in the STT condition, the second and third block were in test format. We also included the

Condition	Presentation 1	Presentation 2	Presentation 3	Final Test
SSS	Study The optic nerve is composed of the axons of ganglion cells.	Study The optic nerve is composed of the axons of ganglion cells.	Study The optic nerve is composed of the axons of ganglion cells.	The optic nerve is composed of the _____ of ganglion cells.
	Study Romanesque and Gothic are two styles of cathedral.	Study Romanesque and Gothic are two styles of cathedral.	Test Romanesque and _____ are two styles of cathedral.	Romanesque and _____ are two styles of cathedral.
STS	Study Robert Frost recited a poem at the inauguration of President Kennedy.	Test Robert Frost recited a poem at the inauguration of President _____.	Study Robert Frost recited a poem at the inauguration of President Kennedy.	Robert Frost recited a poem at the inauguration of President _____.
	Study The Mongol empire occupied Russia for 240 years.	Test The _____ empire occupied Russia for 240 years.	Test The _____ empire occupied Russia for 240 years.	The _____ empire occupied Russia for 240 years.
Never- Studied	no study or self-test			_____ gives Mars its notorious red hue.

Day 1 Day 1 or 2

FIGURE 2 Structure of training conditions and sample stimuli. Full stimulus set is available on the Open Science Framework (see open practices statement at end of manuscript for link).

STS condition, that we compared to the SST condition to test our hypothesis that having the opportunity to restudy material after test (STS) would provide additional long-term memory benefit compared to restudying material before test (SST). The order of presentation of the blocks from the four training conditions was counterbalanced using a Latin Square design, so each condition appeared equally often in all serial positions. Each block was ~5 min in length, making the total training phase time ~60 min. Participants were given opportunities to take self-paced breaks after each block to reduce fatigue.

After completion of the computerized training phase, participants were given a written final test containing 150 test items: 120 statements from training and 30 new statements that were never studied. They were asked to write down (by hand) the missing keyword from memory to accurately complete each statement. Participants randomly assigned to the immediate final test condition (Experiment 1: $n = 31$, $M_{age} = 21.3$, $SD_{age} = 3.3$, age range: 18–31, 21 females, 10 males; Experiment 2: $n = 38$, $M_{age} = 20.3$, $SD_{age} = 2.2$, age range: 18–29, 31 females, 7 males) were given the written test during session one, immediately after completion of the computerized learning task. Participants assigned to the delay final test condition (Experiment 1: $n = 34$, $M_{age} = 22.3$, $SD_{age} = 3.8$, age range: 19–33, 20 females, 14 males; Experiment 2: $n = 33$, $M_{age} = 19.9$, $SD_{age} = 2.5$, age range: 18–32, 23 females, 10 males) completed the final written test during their second session, 2 days later in Experiment 1 or 7 days later in Experiment 2. Although participants assigned to the immediate group had already taken their final written test during the first

session, they were also brought back to the lab for a second session and completed unrelated tasks to avoid a self-selection bias for a two-session vs. single-session study.

Results

The facts that participants endorsed as already known to them ($M = 21\%$, $SD 12\%$) were excluded from subsequent analyses to focus on memory for facts learned within the course of the experiments. All results replicate when memory for all studied facts is analyzed instead. Participants' self-report was a meaningful indicator of their prior knowledge as their final recall of those facts was 80%, compared to 59% recall for facts they did not endorse as already known during study [$t(133) = 18.80$, $p < 0.001$, $d = 1.62$]. Participants learned new information over the course of the experiment beyond their prior knowledge, as they were able to fill-out only 11% of blanks for never-studied items during final test.

Self-reported recall during retrieval practice matched objective performance

To verify that subjects were engaged in retrieval practice, we compared self-reported recall success to objective final recall performance. We limited these analyses to the self-reported recall during the last block of the SST and STT conditions because these self-tests were directly followed by a final test (immediate or

delayed) with no extra learning opportunity. In all retrieval practice blocks that were not followed by additional study opportunities, participants recalled 84% of facts they claimed they “definitely remembered,” 29% of facts they claimed they “maybe remembered,” and 6% of facts they claimed they “did not remember.” When limited to the immediate final test group, performance was similar with participants recalling 86% of the facts they claimed they “definitely remembered,” 27% of the facts they claimed they “maybe remembered,” and 8% of the facts they claimed they “did not remember.” Results confirmed that participants were engaged in the task and their self-reported performance on test trials was a meaningful indicator of their performance.

The effect of repetition vs. retrieval practice

The main question of our study was whether the striking benefit of replacing all repetition with retrieval practice (delayed recall STTT > SSST > SSSS), reported by Roediger and Karpicke (2006a), replicates for briefly presented facts. This was not the case. Mean recall for facts encountered in SSS, SST and STT conditions in each group is presented in Figure 3. The recall scores were submitted to a 3 x 2 x 2 repeated measures ANOVA, with study condition [SSS, SST, STT] as a within-subject factor and group [immediate, delay] and experiment [Experiment 1, Experiment 2] as between-subject factors. Full ANOVA results are reported in Table 1. Although there

was a main effect of experiment, reflecting better recall for the shorter delay period in Experiment 1 ($M_{\text{recall}} = 60\%$, 95% CI [56.7–63.3%]) than Experiment 2 ($M_{\text{recall}} = 54\%$, 95% CI [50.3–56.6%]), experiment factor did not interact with any effect of interest. We thus retained the joint report of the data for brevity, verifying that the same pattern of results is present in each experiment separately.

Of main interest was the condition x group interaction, which was significant ($p < 0.001$). To characterize the interaction, we analyzed the effect of condition in each group (immediate, delayed) separately. The immediate group showed the same pattern as the immediate group in Roediger and Karpicke (2006a), with more study resulting in better recall [SSS > SST > STT, $F(2, 136) = 51.87$, $p < 0.001$, $\eta_p^2 = 0.43$, linear trend $F(1.68) = 84.58$, $p < 0.001$, $\eta_p^2 = 0.55$]. However, this pattern did not reverse in the delayed recall group. There were still reliable differences among conditions [$F(2, 132) = 18.66$, $p < 0.001$, $\eta_p^2 = 0.22$], but the pattern was quadratic rather than related to the amount of self-testing linearly [linear trend $F(1.66) = 0.001$, $p = 0.97$, $\eta_p^2 = 0.00$; quadratic trend $F(1.66) = 41.04$, $p < 0.001$, $\eta_p^2 = 0.38$]. Delayed recall was the highest in the SST condition that included repetition in addition to self-testing ($M_{\text{recall}} = 60\%$) rather than in the condition that maximized restudy (SSS) or maximized self-testing (STT), which both showed ~50% recall. The same quadratic pattern, and reliable delayed recall advantage for the SST condition, was observed in both Experiments (both quadratic trend $F > 16.9$, $p < 0.001$; see Figure 3).

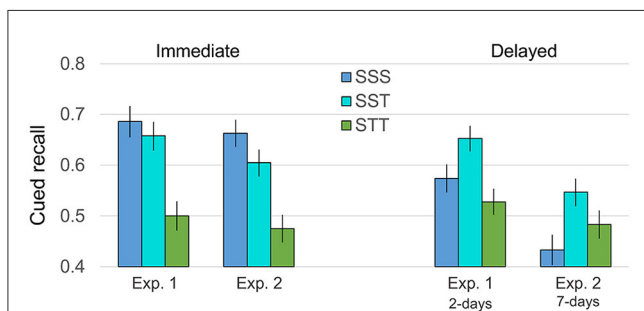


FIGURE 3 Mean cued recall during immediate and delayed fill-in-the-blank recall test comparing SSS, SST, and STT conditions. Delayed test in Experiment 1 came after 2 days while delayed test after Experiment 2 came after 7 days. Error bars denote standard error of the mean.

The effect of restudy before test vs. after test

Our secondary question was whether restudy after test (STS) provides additional benefits compared to a restudy before test (SST). Full results of the 2 x 2 x 2 repeated measures ANOVA, with study condition [SST, STS] as a within-subject factor, and group [immediate, delay] and experiment [Experiment 1, Experiment 2] as between-subject factors are reported in Table 2 and visualized in Figure 4. Of main interest was the effect of study condition and its interactions. The main effect of study condition was significant [$F(1, 132) = 6.19$, $p = 0.014$, $\eta_p^2 = 0.045$], as was the condition x group interaction [$F(1, 132) = 7.60$, $p = 0.007$, $\eta_p^2 = 0.054$].

To follow up on the interaction, we evaluated the effect of test-timing condition (SST vs. STS) within each group. Contrary to our

TABLE 1 Effect of repetition and retrieval practice, repeated measures ANOVA results.

Condition (SSS, SST, STT) x group (immediate, delay) x experiment (Exp 1, Exp 2)				
Effect	df	F	p	η_p^2
Study condition*	2, 264	45.38	<0.001	0.256
Group*	1, 132	7.08	0.009	0.051
Experiment*	1, 132	8.05	0.005	0.057
Study condition x group*	2, 264	27.68	<0.001	0.173
Study condition x experiment	2, 264	2.04	0.133	0.015
Study condition x group x experiment	2, 264	1.80	0.167	0.013

Statistically significant effects are marked by an asterisk. df, degrees of freedom; F, f -statistic; p, p-value; η_p^2 , effect size.

TABLE 2 Effect of restudy before and after test, repeated measures ANOVA results.

Condition (SST, STS) x group (immediate, delay) x experiment (Exp 1, Exp 2)				
Effect	df	F	p	η_p^2
Study condition*	1, 132	6.19	0.014	0.045
Group*	1, 132	6.28	0.013	0.045
Experiment*	1, 132	7.39	0.007	0.053
Study condition x group*	1, 132	7.60	0.007	0.054
Study condition x experiment	1, 132	0.32	0.571	0.002
Study condition x group x experiment	1, 132	0.00	0.997	0.000

Statistically significant effects are marked by an asterisk. df, degrees of freedom; F, f -statistic; p, p-value; η_p^2 , effect size.

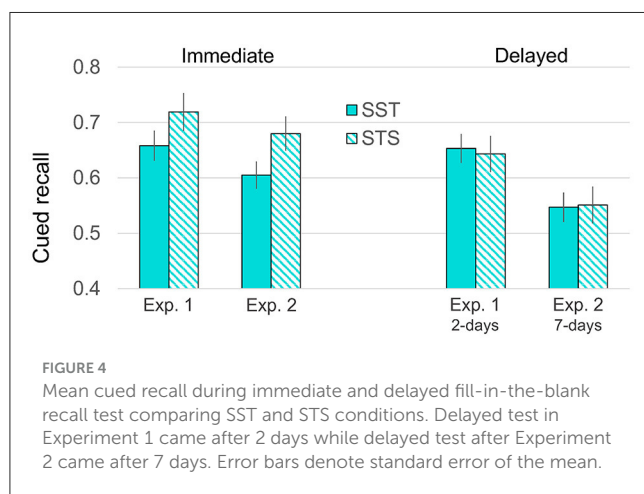


FIGURE 4 Mean cued recall during immediate and delayed fill-in-the-blank recall test comparing SST and STS conditions. Delayed test in Experiment 1 came after 2 days while delayed test after Experiment 2 came after 7 days. Error bars denote standard error of the mean.

prediction, no advantage was found during delayed recall, with 60% recall success in both SST and STS conditions [$t(66) = 0.18, p = 0.86, d = 0.02$]. However, at immediate recall, we found a significant advantage of restudy after self-test, with 70% recall in the STS condition compared to 63% recall in the SST condition [$t(68) = 4.15, p < 0.001, d = 0.50$]. Notably, this advantage of STS condition must have been driven by the second study in this condition. When comparing self-reported retrieval success during the self-test, which was after a single study in the STS condition while after two studies in the SST condition, participants self-reported confident retrieval success on 54% of trials in the STS condition compared to 63% in the SST condition [$t(135) = 7.42, p < 0.001, d = 0.64$]. So, while the retrieval success remained relatively stable from the self-test to the final test in the SST condition, as would be expected with no intervening study opportunity, it improved substantially in the STS condition after the second study.

Retrieval practice improves subjective prediction of encoding success

Our final, exploratory question was whether testing may be beneficial for memory and retention because it improves insight into how well-information has been learned. Specifically, testing may improve self-awareness of one's level of knowledge

allowing for better attentional allocation toward learning on subsequent trials. To examine the effect of testing on how well the participants were able to predict their encoding success during study blocks, we examined the prospective memory judgement ratings that were collected throughout learning on the study trials. First, we wanted to verify that participants' prospective memory judgements ("will remember," "may remember") during study trials were meaningfully related to their actual subsequent memory performance at final test. Indeed, across all conditions, we found better memory performance at final test for items confidently predicted to be remembered in the future ($M_{acc} = 76\%, SD_{acc} = 16\%$) than for items that were not confidently endorsed [$M_{acc} = 48\%, SD_{acc} = 17\%; t(136) = 23.24, p < 0.001, d = 1.985$]. Thus, metacognitive judgements indicated that participants were learning information during the experiment and developing a good sense of how well they were learning the material throughout.

Next, we were interested in whether testing improved participants' metamemory, measured as the difference in final test accuracy for high confidence ("will remember") trials and low confidence ("may remember") trials during the final study block in each condition. For example, a participant that recalled 80% of facts endorsed as "will remember" and 60% of facts endorsed as "may remember" would be considered to have higher ability to predict encoding success (i.e., better metamemory) than a subject who recalls 70% of facts endorsed as "will remember" and 70% of facts endorsed as "may remember." To determine whether metamemory is enhanced by an opportunity to restudy following testing, we compared subjective predictions of encoding success between the STS and SST conditions: the STS condition was the only condition that allowed a restudy opportunity during the final block after self-testing and the SST condition was the only condition that also allowed a restudy opportunity during the final block but with no self-testing opportunities. Differences in metamemory between these conditions provide additional information about how testing prior to restudy may impact memory. Metamemory scores were submitted to a $2 \times 2 \times 2$ repeated measures ANOVA with study condition [SST, STS] as a within-subject factor and group [immediate, delay] and experiment [Experiment 1, Experiment 2] as a between-subjects factors, as reported in Table 3. There was no main effect of experiment and no interactions. However, the main effect of study condition was significant [$F(1, 108) = 16.91, p < 0.001, \eta_p^2 = 0.14$], indicating subjects were more accurate at predicting their encoding success in the STS condition compared

TABLE 3 Effect of restudy before test on predicted encoding success, repeated measures ANOVA results.

Condition (STS, SSS) x group (immediate, delay) x experiment (Exp 1, Exp 2)				
Effect	df	F	p	η_p^2
Study condition*	1, 108	16.91	<0.001	0.135
Group*	1, 108	7.68	0.007	0.066
Experiment	1, 108	0.80	0.374	0.007
Study condition x group	1, 108	0.01	0.935	0.000
Study condition x experiment	1, 108	0.31	0.580	0.003
Study condition x group x experiment	1, 108	1.22	0.271	0.011

Statistically significant effects are marked by an asterisk. df, degrees of freedom; F, f-statistic; p, p-value; η_p^2 , effect size.

to the SSS condition ($M_{\text{diff}} = 0.12$, $SE_{\text{diff}} = 0.03$, $p < 0.001$, 95%CI [0.07, 0.18]). Together, these results suggest that one way testing improves memory is by increasing awareness of not-yet-mastered material, which may make subsequent restudy more efficient.

Interestingly, the main effect of group was also significant [$F(1, 108) = 7.68$, $p = 0.007$, $\eta_p^2 = 0.07$] with individuals in the delay group showing overall greater correspondence between predicted memory and the actual final test success (accuracy difference between “will remember” and “may remember” facts, $M = 0.48$, $SE = 0.02$) than individuals in the immediate group (accuracy difference between “will remember” and “may remember” facts, $M = 0.39$, $SE = 0.02$). The group differences were primarily driven by the low confidence “may remember” judgments: participants in the immediate group recalled 84% of facts they previously endorsed as “will remember” and 45% of facts they endorsed as “may remember” on the last study. In contrast, after delay, participants still recalled 73% of the facts they as endorsed as “will remember” but only 25% of facts they endorsed as “may remember” on the last study. Thus, the “may remember” facts were more likely to be forgotten over time.

Discussion

A classic study by Roediger and Karpicke (2006a) demonstrated a striking benefit of replacing study repetition with retrieval attempts when studying a text passage; 1 week after a study session, recall was highest in participants who had only a single study period followed by multiple retrieval periods rather than participants afforded additional study time. Using briefly presented academically relevant facts, the current study showed a different pattern of results. Specifically, we did not find evidence that individuals were better off by replacing all study repetitions with retrieval attempts. Instead, our results showed that including a study repetition in addition to self-testing (SST) lead to the highest delayed recall, outperforming both all-study (SSS) and multi-test (STT) conditions. Our secondary question was whether having an opportunity to restudy after a test rather than before a test may lead to better learning and retention, as tests have been shown to enhance subsequent learning. The results were mixed in that regard. Participants indeed better remembered items that were restudied after a test (STS) than items that were restudied before a test (SST), but only when tested immediately. Delayed recalled

was equivalent across STS and SST conditions, indicating that the retrieval attempt *per se* provided a comparable benefit to memory, irrespective of timing.

Our results emphasizing the importance of repetition in addition to retrieval practice are intuitive, but contrast with the results of the seminal Roediger and Karpicke (2006a) study that inspired our task design. In Experiment 2 of Roediger and Karpicke (2006a), repetition led to superior initial learning (SSSS > SSST > STTT immediate recall), but the order of conditions flipped completely for delayed recall (SSSS < SSST < STTT recall after a week delay). As each study period was minutes long and contained repeated readings of the material within a single study period, here we asked whether the same pattern of results replicates when each study opportunity is more time limited. In our study, the benefit of retrieval practice on preventing forgetting was also observed, as evidenced by diminishing differences between immediate and delayed recall performance in conditions that included self-testing. However, the self-testing benefits did not outweigh the benefit of repetition. Why does repetition seem more important in our study? Notably, a single “S” in past studies rarely meant a single brief presentation; rather, additional study repetition was often included for all items. For example, in Roediger and Karpicke (2006a) participants in the single-study STTT condition studied prose passages for a fixed 5-min period, which provided them with an average of 3.4 study repetitions before testing. In Karpicke and Roediger (2007) a single study block consisted of five study repetitions of each wordlist. Thus, our findings of the importance of repetition are not at odds with these seminal papers. Rather, our results make the well-known importance of repetition for long-term retention (Davis et al., 1972; Donaldson, 1981; Mayer, 1983; Bromage and Mayer, 1986; Webb, 2007) more explicit in the context of retrieval-practice effects. Balancing testing with repetition may be optimal for maximizing retention in the current study because it combined the benefits of study repetition for learning and retrieval practice for retention.

A second goal of our study was to further our understanding of test-potentiated learning by comparing the relative benefits of restudy before retrieval practice vs. restudy after retrieval practice (Izawa, 1966, 1970; Arnold and McDermott, 2013). Given prior work that showed retrieval attempts, even when unsuccessful, made subsequent study more productive (Kornell et al., 2009; Kornell, 2014), we hypothesized that restudy after test (STS) would boost learning and long-term retention to a greater degree

than restudying material before testing (SST). Importantly, only the STS condition provided an opportunity for test-potentiated learning. Because no feedback was provided on retrieval attempts themselves, we avoided potentially conflating retrieval benefits with test-potentiated learning in all other conditions (Arnold and McDermott, 2013). For the immediate test, the results were consistent with our hypothesis: final test recall was greater in the STS than the SST condition, and more accurate prediction of retrieval success at the final study for the STS than for the SSS condition. Self-testing may provide participants with the internal feedback that they have not yet mastered the material to the level necessary to support recall, thus helping to guide attention and make subsequent study more fruitful. Restudying after test may also confer benefits by increasing spacing between the two study opportunities, as past work has shown that spacing repetitions further apart results in better recall than when items are studied *en masse* (Melton, 1970; Hintzman and Rogers, 1973; Cepeda et al., 2006; Kornell and Bjork, 2008). For example, in the current study, study repetitions in the SSS and SST conditions were separated by 3 blocks of intervening material while study repetitions in the STS condition was separated by 7 blocks of intervening material.

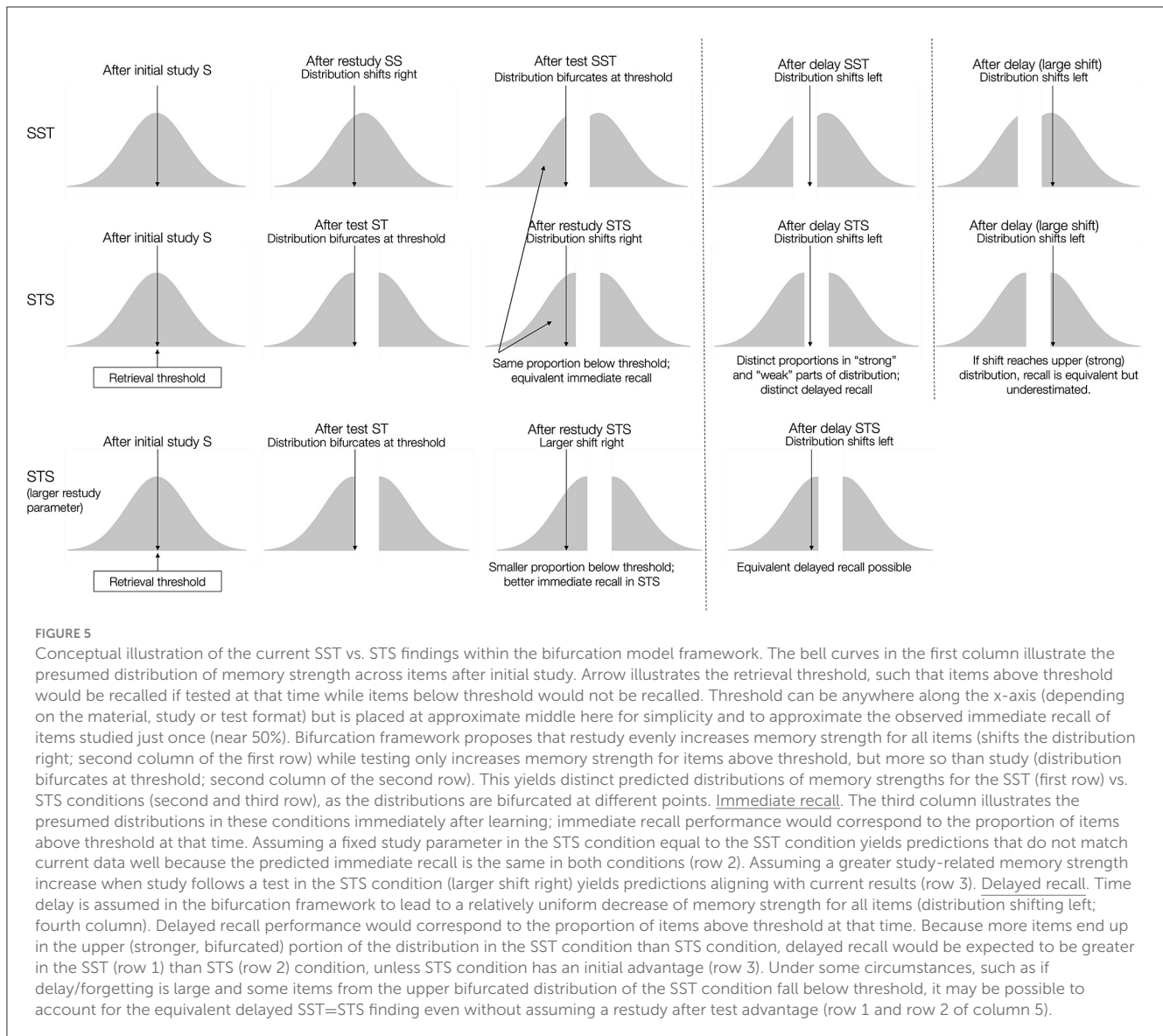
Curiously, the benefit of restudying after test rather than before test was no longer observed when the final test came on a later day, even though the benefit of a study after test (Kornell, 2014) or the benefit of spacing study repetitions apart (Bahrick et al., 1993; Cepeda et al., 2008) has been observed even when a test is administered on a different day. Yet, here we observed equivalent long-term memory in the STS and SST conditions, and replicated this finding across two experiments with different delays. We speculate that while less was learned initially in the SST compared to STS condition, what was learned in the SST condition was easier to retain as more material was reinforced by the self-test prior to final test (see a detailed discussion in the bifurcation framework below). This is a novel observation that contrasts with prior work that manipulated study and test order and found better delayed memory for interleaved conditions (e.g., STST > SSTT in Karpicke and Roediger, 2007). As prior studies typically provided a final retrieval practice in both conditions, this additional opportunity to reinforce restudied items may have provided an additional boost to performance in the interleaved condition by combining the strength of test-potentiated learning effects and retrieval practice effects (Arnold and McDermott, 2013). Without the additional retrieval practice opportunity, we isolated test-potentiated learning effects from retrieval practice effects to a greater degree than prior work, finding that restudy after test and restudy before test may be equally beneficial for long-term retention (delayed STS = SST).

The mechanisms through which retrieval practice impacts long-term memory retention are still a matter of debate. One possibility is that retrieval practice provides additional pathways through which to reactivate knowledge (Collins and Loftus, 1975; Anderson, 1983). According to this viewpoint, repetition serves to reactivate the same memory exactly as it was encoded and therefore strengthens the connections from the originally encoded pathway for retrieval. But testing, especially cued- or free-recall, may activate other related concepts and provide additional pathways through which one can retrieve the same information (Carpenter, 2009,

2011; Pyc and Rawson, 2010). Our finding of maximal long-term retention in the STS and SST conditions (with no difference between them) is consistent with this perspective as participants had the opportunity to *both* strengthen the original pathway through repetition *and* create a new pathway through cued-recall testing in both conditions. A recent study by Zheng et al. (2016) further supported the idea of multiple pathways by demonstrating that the benefit of testing can be pushed even further, and memory retention increased, when testing involves practicing multiple retrieval routes as opposed to just one. This is consistent with older work that has shown that varying the format of test or the retrieval strategy employed across repeated testing provides additional benefits (Anderson and Pichert, 1978; Gilbert and Fisher, 2006; Finley, 2012). While a balance between study and testing led to the best long-term memory retention outcome in the current paradigm, it may be interesting to probe in future studies whether a condition with multiple tests could become superior if each test had a different format, potentially providing more pathways to the same information.

The benefits of testing are typically viewed in terms of protection against forgetting that is qualitatively different from repetition alone (Nungester and Duchastel, 1982; Roediger and Karpicke, 2006b; Roediger and Butler, 2011). This view is consistent with neuroscience of memory consolidation that has shown distinct neural substrates for consolidated memories that make them stable and resistant to forgetting (Huang, 1998). Furthermore, retrieval elicited by testing can trigger rapid consolidation resulting in memory stabilization (Antony et al., 2017); see also Wing et al. (2013) and Keresztes et al. (2014). However, Kornell et al. (2011) noted that one does not need to postulate qualitative memory differences and distinct forgetting rates between repeated vs. tested items. Instead, they propose that memory strength of all items may be falling at a constant rate and the apparent lack of forgetting after retrieval results from a *bifurcation*—or split—of the distribution of memory strength values in response to retrieval. A retrieval attempt (or testing without feedback) strengthens a subset of items that are above threshold and can be recalled at the moment but does not strengthen items that are below threshold and thus not recalled. This results in a distribution of memory strengths that is bifurcated, with a subset that is below threshold and another that is high above threshold. Restudy, on the other hand, strengthens all items but less so than testing. Thus, even if memory strength for both restudied items and tested items falls at the same rate, it takes the retrieval-boosted items much longer to reach the threshold below which they are no longer retrievable, leading to an apparent lack of forgetting. We suspect that retrieval practice results in an actual reduction in forgetting thanks to structural changes in the neural substrate (Huang, 1998). Nevertheless, either framing of retrieval effects—decreased forgetting rates or selective disproportionate increases of memory strength—are equally applicable to the current behavioral data.

Finally, the current findings of STS advantage over SST at immediate test but not after a delay inform current models of testing effects and support the notion that restudy after a test yields greater benefits than restudy before a test (Kornell, 2014). For example, using the bifurcation model framework



(Kornell et al., 2011), let us consider what would be predicted in each condition if one assumes (a) a single study parameter or (b) a higher study parameter for restudy after test. As schematically illustrated in Figure 5, assuming a single study parameter that is equivalent between SST condition (first row) and STS condition (second row) does not predict our current data well. Whether the distribution of memory strengths first shifts right and then is bifurcated at threshold in the SST condition, or is first bifurcated and then shifts right in the STS condition, the same number of items remain below threshold immediately after learning, and similar performance would be expected at immediate test. After a delay, one would expect a greater benefit for the SST condition because more items are in the stronger, bifurcated part of the distribution and thus less likely to reach forgetting threshold as memory strength for all items decreases over time. Neither of these predictions matches the current data.

In contrast, assuming that study-related increase in memory strength is greater when restudy follows testing in the STS

condition, one would expect immediate recall benefit for the STS condition, with fewer items below threshold (Figure 5, third row). After a delay, relatively equal performance can be also accounted for: fewer items would reach forgetting threshold in the SST condition than in the STS condition because SST condition yields more items in the stronger portion of the bifurcated distribution. Thus, assuming a larger shift in memory strength for restudy after test better accounts for the current data than assuming a fixed study benefit.

Another piece of evidence that restudy is more efficient after a self-test comes from the analysis examining subjective reports of encoding success. Specifically, participants' predictions about their fact memory were more closely differentiating actually remembered and actually forgotten facts when they had an opportunity to self-test rather than just re-read the facts. This may allow them to efficiently allocate more study effort to not-yet-mastered items (Nelson and Dunlosky, 1991; Metcalfe and Kornell, 2005; Metcalfe and Finn, 2008a; Little and McDaniel,

2015). Importantly, the relative benefits of restudy after self-test (STS; presumably improving restudy efficiency) vs. self-testing after restudy (SST; presumably strengthening all facts learned across either repetition) may be context dependent. Although we found about equal delayed performance after both a 2-day delay and 7-day delay, the bifurcation framework would predict that the relative benefit may depend on the length of the delay, given that the distributions are bifurcated at different points. Future studies may test this prediction by using a greater range of delays.

Conclusions

Understanding how testing should be balanced with opportunities to restudy material is important for determining optimal learning strategies that can be implemented in personal study and inform educational approaches. Here, we found that when study opportunities are limited to a single brief presentation, akin to briefly encountering information in an academic lecture or textbook, balancing self-testing with repetition may provide the maximal benefit. Restudying after a retrieval attempt provided an additional benefit compared to restudying before a retrieval attempt on an immediate test, but this benefit did not carry over a delay. These results provide new constraints for existing cognitive theories of testing effects, offer new insights into the conditions that promote encoding and long-term retention, and have practical implications for education.

Open practices

None of the experiments discussed in the current report were preregistered. Data and stimuli are freely available in the *A Combination of Restudy and Retrieval Practice Maximizes Retention of Briefly Encountered Facts* repository on the Open Science Framework (<https://osf.io/3fkup/>). A formal power analysis with an effect size estimate unique to the current study was not performed. Rather, we used a target sample size that we previously determined to be sufficient for detecting medium-sized effects.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: “A Combination of Restudy and Retrieval Practice Maximizes Retention of Briefly Encountered

Facts” repository on the Open Science Framework (<https://osf.io/3fkup/>).

Ethics statement

The studies involving humans were approved by Research Compliance Services at the University of Oregon. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

SA: Conceptualization, Formal analysis, Methodology, Project administration, Visualization, Writing—original draft, Writing—review and editing, Investigation. DZ: Conceptualization, Investigation, Methodology, Resources, Supervision, Writing—original draft, Writing—review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author DZ declared that they were an editorial board member of *Frontiers* at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Adesope, O. O., Trevisan, D. A., and Sundararajan, N. (2017). Rethinking the use of tests: a meta-analysis of practice testing. *Rev. Edu. Res.* 87, 659–701. doi: 10.3102/0034654316689306
- Anderson, J. R. (1983). A spreading activation theory of memory. *J. Verb. Learn. Verb. Behav.* 22, 1–33.
- Anderson, R. C., and Pichert, J. W. (1978). Recall of previously unrecalled information following a shift in perspective. *J. Verb. Learn. Verb. Behav.* 17, 1–12. doi: 10.1016/S0022-5371(78)90485-1
- Antony, J. W., Ferreira, C. S., Norman, K. A., and Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends Cogn. Sci.* 21, 573–576. doi: 10.1016/j.tics.2017.05.001
- Arnold, K. M., and McDermott, K. B. (2013). Test-potentiated learning: distinguishing between direct and indirect effects of tests. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 940–945. doi: 10.1037/a0029199
- Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., and Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychol. Sci.* 4, 316–321.

- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vis.* 10, 433–436. doi: 10.1163/156856897X00357
- Bromage, B. K., and Mayer, R. E. (1986). Quantitative and qualitative effects of repetition on learning from technical text. *J. Edu. Psychol.* 78, 271–278. doi: 10.1037/0022-0663.78.4.271
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1563–1569. doi: 10.1037/a0017021
- Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: support for the mediator effectiveness hypothesis of the testing effect. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1547–1552. doi: 10.1037/a0024140
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., and Rohrer, D. (2006). Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychol. Bull.* 132, 354–380. doi: 10.1037/0033-132.3.354
- Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., and Pashler, H. (2008). Spacing effects in learning: a temporal ridge of optimal retention. *Psychol. Sci.* 19, 1095–1102. doi: 10.1111/j.1467-9280.2008.02209.x
- Collins, A. M., and Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychol. Rev.* 82, 407–428. doi: 10.1037/0033-295X.82.6.407
- Davis, J. C., Lockhart, R. S., and Thomson, D. M. (1972). Repetition and context effects in recognition memory. *J. Exp. Psychol.* 92, 96–102. doi: 10.1037/h0032157
- Dirkx, K. J. H., Kester, L., and Kirschner, P. A. (2014). The testing effect for learning principles and procedures from texts. *J. Edu. Res.* 107, 357–364. doi: 10.1080/0022013823370
- Donaldson, W. (1981). Context and repetition effects in recognition memory. *Mem. Cogn.* 9, 308–316. doi: 10.3758/BF03196964
- Ebbinghaus, H. (1913). *Memory: A Contribution to Experimental Psychology*, eds H. A. Ruger and C. E. Bussenius. New York, NY: Teachers College Press.
- Finley, J. R. (2012). *Retrieval cue variability: When and Why are Two Meanings Better than One?* (Doctoral dissertation). University of Illinois at Urbana, Champaign, IL, United States.
- Gates, A. I. (1917). *Recitation As a Factor In Memorizing*. Columbia University.
- Gilbert, J. A. E., and Fisher, R. P. (2006). The effects of varied retrieval cues on reminiscence in eyewitness memory. *Appl. Cognit. Psychol.* 20, 723–739. doi: 10.1002/acp.1232
- Greving, S., and Richter, T. (2018). Examining the testing effect in university teaching: retrievability and question format matter. *Front. Psychol.* 9, 1–10. doi: 10.3389/fpsyg.2018.02412
- Hays, M. J., Kornell, N., and Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 290–296. doi: 10.1037/a0028468
- Hintzman, D. L., and Rogers, M. K. (1973). Spacing effects in picture memory. *Mem. Cogn.* 1, 430–434. doi: 10.3758/BF03208903
- Hogan, R. M., and Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *J. Verb. Learn. Verb. Behav.* 10, 562–567. doi: 10.1115/DSCC2008-2224
- Huang, E. P. (1998). Synaptic plasticity: going through phases with LTP. *Curr. Biol.* doi: 10.1016/s0960-9822(98)70219-2
- Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychol. Rep.* 18, 879–919. doi: 10.2466/pr0.1966.18.3.879
- Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *J. Exp. Psychol.* 83(2 PART 1), 340–344. doi: 10.1037/h0028541
- Jonsson, F. U., Kubik, V., Sundqvist, M. L., Todorov, I., and Jonsson, B. (2014). How crucial is the response format for the testing effect? *Psychol. Res.* 78, 623–633. doi: 10.1007/s00426-013-0522-8
- Kao, Y. C., Davis, E. S., and Gabrieli, J. D. E. (2005). Neural correlates of actual and predicted memory formation. *Nat. Neurosci.* 8, 1776–1783. doi: 10.1038/nn1595
- Karpicke, J. D., and Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *J. Mem. Lang.* 57, 151–162. doi: 10.1016/j.jml.09004
- Keresztes, A., Kaiser, D., Kovcs, G., and Racsmy, M. (2014). Testing promotes long-term learning via stabilizing activation patterns in a large network of brain areas. *Cereb. Cort.* 24, 3025–3035. doi: 10.1093/cercor/bht158
- Koriat, A., Sheffer, L., and Ma'ayan, H. (2002). Comparing objective and subjective learning curves: judgments of learning exhibit increased underconfidence with practice. *J. Exp. Psychol. Gen.* 131, 147–162. doi: 10.1037/0096-3445.131.2.147
- Kornell, N. (2014). Attempting to answer a meaningful question enhances subsequent learning even when feedback is delayed. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 106–114. doi: 10.1037/a0033699
- Kornell, N., and Bjork, R. A. (2008). Learning concepts and categories: is spacing the “enemy of induction”? *Psychol. Sci.* 19, 585–592. doi: 10.1111/j.1467-9280.2008.02127.x
- Kornell, N., Bjork, R. A., and Garcia, M. A. (2011). Why tests appear to prevent forgetting: a distribution-based bifurcation model. *J. Mem. Lang.* 65, 85–97. doi: 10.1016/j.jml.04002
- Kornell, N., Hays, M. J., and Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 989–998. doi: 10.1037/a0015729
- Little, J. L., and McDaniel, M. A. (2015). Metamemory monitoring and control following retrieval practice for text. *Mem. Cogn.* 43, 85–98. doi: 10.3758/s13421-014-0453-7
- Lovelace, E. A. (1984). Metamemory: monitoring future recallability during study. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 756–766.
- Mayer, R. E. (1983). Can you repeat that? Qualitative effects of repetition and advance organizers on learning from science prose. *J. Edu. Psychol.* 75, 40–49. doi: 10.1037/0022-0663.75.1.40
- Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *J. Verb. Learn. Verb. Behav.* 9, 596–606. doi: 10.1016/S0022-5371(70)80107-4
- Metcalf, J., and Finn, B. (2008a). Evidence that judgments of learning are causally related to study choice. *Psychon. Bull. Rev.* 15, 174–179. doi: 10.3758/PBR.15.1.174
- Metcalf, J., and Finn, B. (2008b). Familiarity and retrieval processes in delayed judgments of learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 1084–1097. doi: 10.1037/a0012580
- Metcalf, J., and Kornell, N. (2005). A Region of Proximal Learning model of study time allocation. *J. Mem. Lang.* 52, 463–477. doi: 10.1016/j.jml.12001
- Nelson, T. O., and Dunlosky, J. (1991). When people’s judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: the “Delayed-JOL Effect.” *Psychol. Sci.* 2, 267–270.
- Nelson, T. O., and Narens, L. (1994). “Why investigate metacognition?” in *Metacognition: Knowing About Knowing*, eds Metcalfe and A. P. Shimamura (MIT Press), 1–25.
- Nungester, R. J., and Duchastel, P. C. (1982). Testing vs. review: effects on retention. *J. Edu. Psychol.* 74, 18–22. doi: 10.1037/0022-0663.74.1.18
- Putnam, A. L., and Roediger, H. L. (2013). Does response mode affect amount recalled or the magnitude of the testing effect? *Mem. Cogn.* 41, 36–48. doi: 10.3758/s13421-012-0245-x
- Pyc, M. A., and Rawson, K. A. (2010). Why testing improves memory: mediator effectiveness hypothesis. *Science* 330, 335. doi: 10.1126/science.1191465
- Richland, L. E., Kornell, N., and Kao, L. S. (2009). The pretesting effect: do unsuccessful retrieval attempts enhance learning? *J. Exp. Psychol. Appl.* 15, 243–257. doi: 10.1037/a0016496
- Roediger, H. L., and Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends Cogn. Sci.* 15, 20–27. doi: 10.1016/j.tics.09003
- Roediger, H. L., and Karpicke, J. D. (2006a). Test-enhanced learning: taking memory tests improves long-term retention. *Psychol. Sci.* 17, 249–255. doi: 10.1111/j.1467-9280.2006.01693.x
- Roediger, H. L., and Karpicke, J. D. (2006b). The power of testing memory: basic research and implications for educational practice. *Perspect. Psychol. Sci.* 1, 181–210. doi: 10.1111/j.1745-6916.2006.00012.x
- Rowland, C. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychol. Bull.* 140, 1–32. doi: 10.1037/a0037559
- Smith, M. A., Roediger, H. L., and Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 1712–1725. doi: 10.1037/a0033569
- Spitzer, H. F. (1939). Studies in retention. *J. Edu. Psychol.* 30, 641–656. doi: 10.1037/h0063404
- Sundqvist, M. L., Mäntylä, T., and Jönsson, F. U. (2017). Assessing boundary conditions of the testing effect: on the relative efficacy of covert vs. overt retrieval. *Front. Psychol.* 8, 1–15. doi: 10.3389/fpsyg.2017.01018
- Tauber, S. K., Witherby, A. E., Dunlosky, J., Rawson, K. A., Putnam, A. L., College, C., et al. (2018). Does covert retrieval benefit learning of key-term definitions? *J. Appl. Res. Mem. Cogn.* 7, 106–115. doi: 10.1016/j.jarmac.10004
- Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *J. Verb. Learn. Verb. Behav.* 6, 175–184.
- van Gog, T., and Kester, L. (2012). A test of the testing effect: acquiring problem-solving skills from worked examples. *Cogn. Sci.* 36, 1532–1541. doi: 10.1111/cogs.12002
- Verkoeijen, P. P. J. L., Bouwmeester, S., and Camp, G. (2012). A short-term testing effect in cross-language recognition. *Psychol. Sci.* 23, 567–571. doi: 10.1177/0956797611435132
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Appl. Ling.* 28, 46–65. doi: 10.1093/applin/aml048
- Wing, E. A., Marsh, E. J., and Cabeza, R. (2013). Neural correlates of retrieval-based memory enhancement: an fMRI study of the testing effect. *Neuropsychologia* 51, 2360–2370. doi: 10.1016/j.neuropsychologia.04004
- Zheng, J., Zhang, W., Li, T., Liu, Z., and Luo, L. (2016). Practicing more retrieval routes leads to greater memory retention. *Acta Psychol.* 169, 109–118. doi: 10.1016/j.actpsy.05014