Memory & Cognition (2022) 50:1299–1318
https://doi.org/10.3758/s13421-022-01324-z

# Exploring the relationship between retrieval practice, self-efficacy, and memory

Andrea N. Frankenstein[1] · Onyinye J. Udeogu[1] · Matthew P. McCurdy[1] · Allison M. Sklenar[1] · Eric D. Leshikar[1]

Accepted: 13 May 2022 / Published online: 6 June 2022


## Abstract

Retrieval practice effect refers to improved memory on a final test for information retrieved one or more times. Although past theoretical work identifies cognitive mechanisms to explain retrieval practice benefits, it is possible that improving self-efficacy during learning may also contribute to better memory, in line with limited past work showing a relationship between self-efficacy and memory. Across two experiments, we examine the potential relationship between retrieval practice, self-efficacy, and memory. In Experiment 1, we examined the extent change in self-efficacy accounted for improved memory on a final test after retrieval practice compared with restudy. In Experiment 2, we gave participants (false) feedback that was either negative (i.e., you performed worse than others), neutral (i.e., you performed the same as others), or positive (i.e., you performed better than others) to more directly assess the effects of self-efficacy on memory under retrieval practice conditions. Results of Experiment 1 showed a significant retrieval practice effect, with memory on the final test being better after retrieval practice compared with restudy. Self-efficacy did not significantly mediate the retrieval practice effect. Results of Experiment 2, however, showed that decreases in self-efficacy due to (false) negative feedback resulted in worse memory performance compared with neutral feedback. Such findings may suggest that change in self-efficacy after retrieval practice attempts, particularly negative feedback, affects memory at final test. Overall, these findings suggest a relationship between retrieval practice, self-efficacy, and memory, and imply that interventions that influence self-efficacy may be a plausible mechanism to modulate memory under some conditions.

**Keywords** Retrieval practice · Self-efficacy · Memory

Decades of research has focused on discovering methods to improve memory. One such procedure to improve memory is retrieval practice, where memory is improved for information that is retrieved one or more times before a final test compared with information that is simply restudied. In a typical retrieval practice procedure, participants learn information under either retrieval practice or restudy conditions before then taking a final test to measure memory. After initial exposure to material, participants are then reexposed to material either in a retrieval practice attempt (e.g., multiple choice, short answer, free recall), or through restudying (where they simply re-read or review material). The common finding in past experimental work is that retrieval practice leads to improved memory compared with restudy (see Delaney et al., 2010; Roediger & Karpicke, 2006; Rohrer & Pashler, 2010; Rowland, 2014, for reviews) across a range of different types of materials (e.g., word lists, paired associates, prose passages, online statistics lectures; Szpunar et al., 2013; for reviews, see Adesope et al., 2017; Pan & Rickard, 2018; Rowland, 2014). Indeed, past work has shown that retrieval practice is such an effective learning strategy that even content that was not explicitly retrieved in a practice test, but was simply related to retrieved content, can also exhibit memory improvements (e.g., Butler, 2010; McDaniel et al., 2013; Rohrer et al., 2010). These findings have immense practical implications for educational settings. In this investigation, we examine a potential social-cognitive mechanism to modulate memory after retrieval practice attempts: self-efficacy.

Researchers have proposed different mechanisms to account for the retrieval practice effect. Most mechanistic accounts focus on cognitive processes underlying the retrieval practice effect, such as strengthening of

✉ Eric D. Leshikar
leshikar@uic.edu

1 University of Illinois at Chicago, 1007 West Harrison Street (M/C 285), Chicago, IL 60607, USA

semantically related information (elaborative retrieval hypothesis; Carpenter, 2009) or promoting similar processing at both encoding and retrieval (transfer-appropriate processing; Morris et al., 1977). Although these cognitive mechanisms have received significant empirical attention and support, less work has investigated the extent noncognitive mechanisms might also contribute to the memory benefits of retrieval practice. One such social-cognitive construct is self-efficacy. Self-efficacy is an individual's belief in their ability in specific situations or their ability to accomplish tasks in novel, unexpected, or potentially stressful contexts (Bandura, 1977; Schunk, 1985). It is plausible that retrieval practice may improve individuals' self-efficacy since it may provide diagnostic information about whether an individual can accurately answer questions on a final test. Self-efficacy can influence how individuals approach goals and challenges, the degree of effort they put into a task, and their persistence when encountering obstacles. Evidence suggests increased self-efficacy is related to increased effort, persistence, and better performance across many domains (Bandura, 1995, 1997). Based on this past work, it may be that modulating self-efficacy affects effort and persistence during learning and final test, and thus contributes to memory effects under retrieval practice conditions.

Before detailing how self-efficacy may contribute to memory improvement for retrieved materials (i.e., the retrieval practice effect in memory), we describe past work that suggests a relationship between self-efficacy and memory. First, work in older adults shows a relationship between self-efficacy and memory such that individuals who report higher memory self-efficacy (higher belief in their ability to remember information) remember more words in free recall tasks compared with those who report lower memory self-efficacy (Beaudoin & Desrichard, 2011). Second, research in educational settings further supports an association between self-efficacy and memory. For example, meta-analytic work shows that academic self-efficacy is related to both academic persistence and academic performance, where higher levels of academic self-efficacy is associated with better academic outcomes (Multon et al., 1991), such as higher student grade point average (Robbins et al., 2004). Because academic success is often tied to memory performance on educational assessments (such as tests), this further implies a relationship between self-efficacy and memory. These lines of work in educational settings demonstrates that academic self-efficacy influences student motivation, learning, and overall achievement (Betz & Hackett, 1983; Pajares, 1996; Schunk & Pajares, 2002). Third, there is experimental evidence that interventions that modulate self-efficacy can improve memory (Huang & Mayer, 2018). In that investigation, participants were assigned to either a treatment group aimed at increasing self-efficacy or a no-treatment control group, before then completing several academic activities (e.g., practice problems, learning-transfer, final test). Results indicated that participants in the self-efficacy treatment group showed higher post-lesson self-efficacy, and further, performed better on practice problems, learning-transfer tests, and most importantly final test performance (i.e., improved memory), compared with control. This evidence suggests that self-efficacy can be modulated, and further, that such changes in self-efficacy are then linked to memory for course materials. Although this past work has shown that self-efficacy can be improved over time and that self-efficacy manipulations can improve learning, it is important to note that this past work has not focused on retrieval practice. We do so in this investigation.

Because the current study investigates whether changes in self-efficacy are associated with memory modulation, it is important to understand how self-efficacy can be changed. Theoretical work suggests that individuals appraise their sense of self-efficacy using information gained from different sources, some of which are mastery experience, vicarious experience, social persuasion, and physiological or emotional arousal (Bandura, 1977). Most relevant to the current investigation is mastery experience. A mastery experience involves completing a task, which can provide a sense that the goal or task can be successfully completed (i.e., can be mastered). As such, mastery experience is an effective method to increase self-efficacy (Bandura, 1977). We propose that retrieval practice could serve as an opportunity for mastery experience that could influence self-efficacy, because the practice test simulates the final test. Specifically, retrieval practice engages the same mental processes (search through memory) that are required on the final test. Further, given that increased self-efficacy is associated with increased persistence on tasks, it may be that changes in self-efficacy affects performance on a final test. Specifically, if retrieval practice does in fact serve as a mastery experience, then engaging in retrieval practice should modulate self-efficacy. Changes in self-efficacy, in turn, could affect task persistence during learning and on the final test, which would influence memory. We took two approaches in the current investigation to examine the relationship between retrieval practice, self-efficacy, and memory. In Experiment 1, participants engaged in either retrieval practice or restudy conditions (without feedback). In Experiment 2, we used an experimental approach that would give more explicit mastery experience information to participants in the form of false (negative, neutral, or positive) feedback. In both experiments, we examined the extent that changes in self-efficacy were linked to changes in memory performance on a final test.

Across two experiments, we investigate the extent retrieval practice influences self-efficacy, and how this change in self-efficacy in turn modulates memory on a final test. In Experiment 1, we examine the extent self-efficacy increases after engaging in retrieval practice and whether this increase in self-efficacy contributes to better memory as measured on a final test compared with restudy. Because self-efficacy is a domain-specific rather than global construct (e.g., a person can have high self-efficacy in some domains but low self-efficacy in others; Bandura, 2006), we used a self-efficacy questionnaire tailored to the learning material. Specifically, participants in Experiment 1 watched four online lecture videos on research methods and engaged in either retrieval practice or restudy after each video (four cycles), and then completed a cumulative final test over all lecture content 48 hours later. Self-efficacy was measured three times during Session 1: at baseline, halfway through the learning procedure (after half the lecture videos were shown), and at the end of the learning procedure, which allows us to measure change in self-efficacy across the experiment. Participants returned 48 hours later for the final test (Session 2). In Experiment 2, we adjusted our design to allow for stronger causal claims about the relationship between retrieval practice, self-efficacy, and memory. We used a similar approach as in Experiment 1 but added a feedback component after retrieval practice. We reasoned that providing feedback would serve as strong mastery experience information, which would allow us to better understand the relationship between retrieval practice, self-efficacy, and memory. All participants engaged in retrieval practice and were assigned to one of three (false) feedback groups: negative feedback (you performed more poorly than others), neutral feedback (you performed the same as others), or positive feedback (you performed better than others). Our approach to directly manipulate feedback (regardless of their actual performance on the retrieval practice task) will more clearly show the possible connection between retrieval practice, self-efficacy, and memory. In particular, we were interested in the (false) negative feedback condition to examine whether reductions in self-efficacy related to negative feedback would account for poorer memory performance. Overall, results of these experiments will yield better understanding of the relationship between retrieval practice, self-efficacy, and memory, which is a potentially rich area of scholarship.

## Experiment 1

There were four predictions in this experiment. First, we hypothesized that self-efficacy would increase in both the retrieval practice and restudy groups (from baseline to end of Session 1) as participants learned about the topic through watching the lecture videos. Second, we predicted that self-efficacy would increase more in the retrieval practice group compared with the restudy group. This would align with our expectations that retrieval practice serves as a mastery experience influencing self-efficacy and would extend past work on the relationship between self-efficacy and memory (Beaudoin & Desrichard, 2011; Huang & Mayer, 2018). Although we expected to see improved self-efficacy in the retrieval practice group compared with restudy, it is also possible that self-efficacy could decrease in the retrieval practice group, reflecting a "correction" to overconfidence in restudy conditions, which is consistent with past work (see Miller & Geraci, 2014). Third, we predicted a standard retrieval practice effect where retrieval practice leads to better memory as measured by the final test compared with restudy (Adesope et al., 2017; Delaney et al., 2010; Roediger & Karpicke, 2006; Rohrer & Pashler, 2010; Rowland, 2014). Further, because prior work demonstrates that retrieval practice can extend to content that is not tested in a practice test but is related to tested materials (see Adesope et al., 2017; Pan & Rickard, 2018; Rowland, 2014), we also predicted to see a retrieval practice effect for new items on the final test (i.e., items that were not seen during retrieval practice/restudy opportunities). Fourth, we predicted that change in self-efficacy would partially account for the retrieval practice effect (using a mediation analysis), which would suggest that self-efficacy contributes to retrieval practice effects in memory. Further, given our experimental design we also examined whether self-efficacy was related to memory overall (across all participants).

### Method

#### Participants

According to a recent meta-analysis (Adesope et al., 2017), the average effect size of the retrieval practice effect is $g$ = 0.51, which is considered a medium effect. We used a more conservative effect size ($f = 0.1$, which is considered a small effect size) to conduct a power analysis using G*Power (Faul et al., 2007) for this experiment, given that the relationship between self-efficacy and memory has not been as well-studied in past experimental work. This analysis indicated a sample size of 100 participants to achieve a power of .80. The study was approved by the university's institutional review board (IRB), and 134 participants (63% female, Age: $M$ = 19.17, $SD$ = 1.38) from the university's psychology subject pool completed Experiment 1. Approximately 33% of participants identified as Latinx/Hispanic, 32% as Asian or Pacific Islander, 13% as White, 9% as Black or African American, and 13% as other races or multiracial. Each participant was randomly assigned to either the restudy or retrieval practice group. We recruited extra participants in anticipation of possible attrition due to our use of a multiday experimental procedure. Ten participants were removed from analysis because they did not complete the second part of the study,

**Table 1** Research methods lecture video topics and video lengths used in both Experiment 1 and Experiment 2

| | Length | Topics |
|---|---|---|
| Video 1 | 8:59 | Scientific Method<br>Operationalizing Variables<br>Reliability and Validity |
| Video 2 | 9:50 | Research designs overview observational research survey research |
| Video 3 | 9:08 | Correlational research experimental research |
| Video 4 | 8:53 | Random assignment controlling bias |

leaving a total of 124 participants with complete data (62 participants per group: retrieval practice, restudy).

### Materials

**Lecture videos** Lecture videos from an Introduction to Research Methods in Psychology course served as the to-be-learned materials. Specifically, we used four videos that ranged from 8:53 to 9:50 each. These videos cover topics such as operationalization of variables, reliability and validity, main types of research designs, and sources of bias (see Table 1 for full list of topics and video lengths). The order of the videos could not be counterbalanced because the videos were designed to be played in a particular sequence (content in later videos refers to earlier videos). Thus, the videos were presented in the same order for all participants.

**Retrieval practice and final test items** Thirty-two multiple-choice items were created based on the lecture video content. Half of the items were definitional questions, and the other half were application questions. Definitional questions required participants to recognize the correct term, definition, or fact stated in the lecture videos, whereas application questions required participants to apply knowledge learned from the lecture videos to a novel example or scenario. Twenty-four items (six per video) were used during the retrieval practice and restudy opportunities. At final test, all 24 items that were used at restudy/retrieval practice were used in the final test (old items), as well as eight new items (two per video). Sample items can be found in Appendix A.

**Self-efficacy questionnaire** Given that self-efficacy is a domain-specific construct, the self-efficacy questionnaire was tailored to the lecture content (e.g., psychology research methods) in accordance with Bandura's (2006) guidelines. Specifically, participants were asked to rate their current degree of confidence in their ability to complete tasks related to research methods (16 questions; 0 "cannot do at all" to 100 "highly certain can do"). For example, "How confident are you that you can explain at least one limitation of correlational research?" (see Appendix B for full list of items.) The same questionnaire was used at Time 1 (baseline), Time 2 (T2), and Time 3 (T3) measurements as described below. In line with previous work, we measured self-efficacy multiple times (see Huang & Mayer, 2018).

### Procedure

This experiment was a mixed design with retrieval practice as the between-subjects factor and self-efficacy as a repeated-measures factor. There was a retrieval practice group and a restudy group. The experiment was composed of two main phases: encoding and reexposure phase (Session 1) and a final test phase 48 hours later (Session 2). Figure 1 depicts the experimental procedure for Session 1. Upon arrival for Session 1, participants reviewed and signed an informed consent document and filled out a demographics questionnaire. Next, participants were given an overview of the different elements of the experiment, including that they would be tested on the learning material during Session 2, and completed a baseline (T1) measurement of self-efficacy. Following the initial measurement of self-efficacy, participants completed the encoding and reexposure phase, during which they completed the second and third self-efficacy measures (T2–T3). All self-efficacy measures were completed during Session 1. Participants returned 48 hours later for Session 2, in which they completed the final test.[1]

**Encoding and reexposure phase** Participants were randomly assigned to one of two groups: retrieval practice or restudy. Participants in both groups completed four encoding/reexposure blocks, where they watched a research methods video before then engaging in a restudy or retrieval opportunity over the material they just learned about in that video (see Fig. 1). The encoding/reexposure blocks were administered via Qualtrics on desktop computers. Specifically, in the encoding and reexposure phase, participants first watched one video on research methods. Participants were told to carefully watch each video as they would be tested over the material in Session 2 (which makes this an intentional learning task). After each video, participants engaged in a distractor task which took approximately 3–5 minutes to complete (because there were four videos, we used four different distractor tasks: forward digit span, backward digit span, digit-symbol substitution, verbal fluency). After completing the distractor task, participants either engaged in retrieval practice or restudy. In the reexposure phase, participants in both groups were exposed to the same 6 multiple-choice items per video (24 total). For the retrieval practice group, participants selected

[1] Self-efficacy was not measured before the final test (Session 2) because we did not want to expose participants to concepts related to the learning materials before taking the final test.

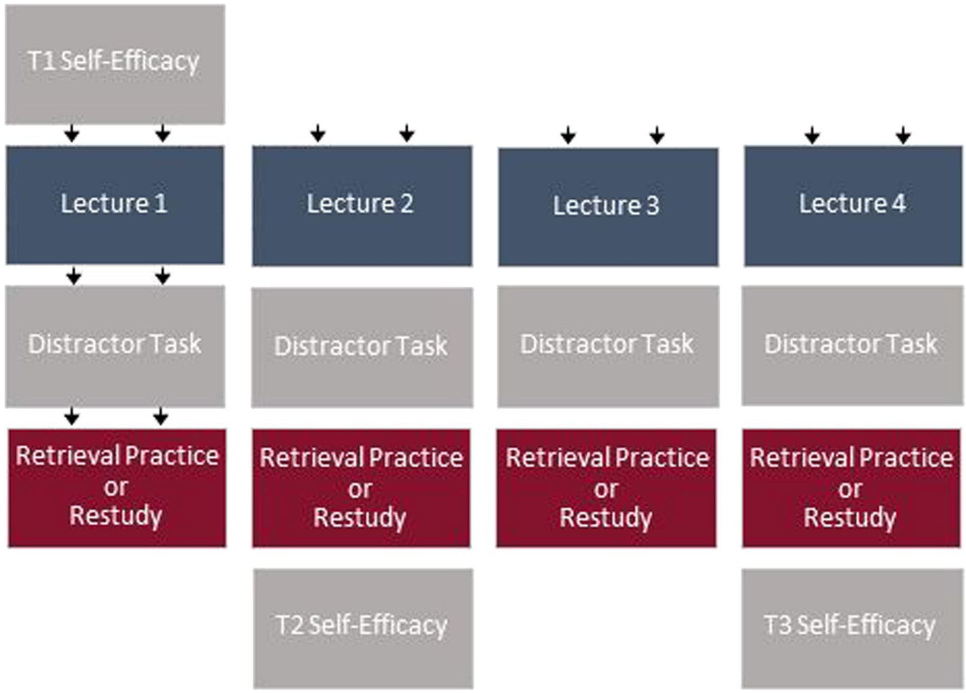


**Fig. 1** Experimental encoding and reexposure procedure for Experiment 1. Participants completed the self-efficacy questionnaire (T1) and the first cycle of lecture video, distractor task, and either retrieval practice or restudy. This cycle repeated for the three subsequent videos, with additional self-efficacy measures after lecture Video 2 (T2) and lecture Video 4 (T3)

**Table 2** Descriptive statistics for distractor tasks by group (restudy, retrieval practice) for Experiment 1

| | Digit Span Total | Digit-symbol Substitution Total | F-A-S Fluency Total |
|---|---|---|---|
| Group | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) |
| Restudy group | 18.45 (4.07) | 61.44 (11.70) | 37.98 (9.63) |
| Retrieval practice group | 19.02 (3.07) | 63.73 (9.69) | 38.64 (9.46) |

Scores did not significantly differ by groups, *ps* > .23

their answer for each item (i.e., multiple-choice item) and were immediately shown the correct answer (no other feedback was given aside from the correct answer). Participants in the restudy group were shown the same multiple-choice questions but with the answers already indicated in bold-face type. Participants in both groups could answer or re-study the items at their own pace but were required to spend a minimum of 7 seconds per item. After completing all tasks associated with the second and fourth videos, both groups rated their self-efficacy using the same self-efficacy questionnaire, which served as the T2 and T3 measures, respectively. At the end of Session 1, participants were dismissed.

**Distractor tasks** Participants completed four distractor tasks on paper, one after each lecture video: forward digit span, backward digit span, digit-symbol substitution (WAIS-R; Wechsler, 1981), and phonemic F-A-S fluency task (Spreen & Benton, 1977; see Appendix C for more details). In addition to serving as a distractor, these measures also allowed us to verify that the retrieval practice group did not differ from the restudy group on measures of fluid intelligence, which could obscure our ability to find a relationship between self-efficacy, retrieval practice, and memory. See Table 2 for descriptive statistics for these tasks as a function of group. Independent t-tests showed that scores on the tasks were not significantly different for the restudy and retrieval practice groups, *ps* > .23. This indicates the two groups did not differ in fluid measures of intelligence, and thus should be comparable.

**Final test phase** After a 48-hour delay, participants returned to the lab to complete a self-paced, pencil-and-paper final cumulative test. Participants were given sheets of paper with all 32 multiple choice questions and were instructed to identify the correct answer for each question. The items on the final test covered materials from all four lecture videos. As described above, 75% of the items were previously seen by participants during reexposure while 25% were new items (i.e., not seen during reexposure).

**Data analysis** All analyses were conducted using the statistical computing software R. We used the *psych* package (Revelle, 2020) to calculate Cronbach's alpha for the self-efficacy measure, *lme4* package (Bates et al., 2015) for the mixed-effects model, and *mediation* package (Tingley et al., 2014) for the mediation analysis.

## Results

In this section we first describe how we scored self-efficacy ratings and performance on the final test, and then report analyses for our four main hypotheses. Self-efficacy was scored by summing responses (0–100) for all items on the self-efficacy questionnaire (maximum score of 1,600), in line with Bandura's (2006) guidelines. The internal consistency of these items, measured by Cronbach's $\alpha$ was .94, which is above the standard .70 cutoff typically considered a stable measure (DeVellis, 2012).[2] We also calculated a self-efficacy change score to assess how self-efficacy changed over the course of the experiment, which served as our primary self-efficacy measure. Self-efficacy change score was calculated by subtracting baseline self-efficacy from final self-efficacy scores (T3 − T1).

Final test performance was scored as the percentage of test items answered correctly (out of 32). We also calculated performance for the previously seen items (i.e., items seen during reexposure) as well as the new items (i.e., items not seen during reexposure), separately. Performance on the previously seen items was calculated as the percentage correct out of 24 (e.g., the number of previously seen items used in this experiment), whereas performance for the new items was calculated as the percentage correct out of 8.

We report self-efficacy as a function of time (T1, T2, T3) and group (retrieval practice, restudy) in Fig. 2. To test the first and second hypotheses (that self-efficacy would increase over time and that this increase would be larger for the retrieval practice group), we conducted a mixed-effects regression. We tested a mixed-effects model with self-efficacy score as the dependent variable, and group (between-subjects factor: restudy, retrieval practice) and time (repeated-measure: T1, T2, and T3) as fixed predictors. The model included each subject's slope of change over time as a random factor to account for individual differences in how participants' self-efficacy changes across Session 1.

[2] Reliability analysis was conducted using all 124 participants' self-efficacy ratings at T1 (baseline). Reliability at T2 and T3 were similar, Cronbach's $\alpha = .94$ and .96, respectively.

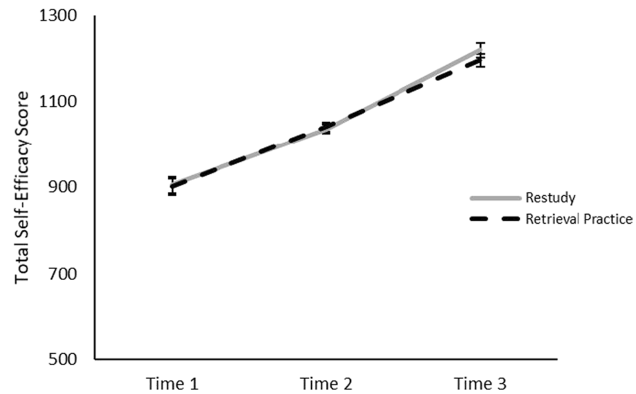


**Fig. 2** Total self-efficacy scores as a function of group (restudy, retrieval practice) and time (T1, T2, T3) for Experiment 1. Error bars represent standard error of the mean. Self-efficacy increased significantly over time, but these increases did not differ as a function of group

**Table 3** Summary of mixed-effects model assessing change in self-efficacy for Experiment 1

| Fixed effects | Estimate | Std. Error |
|---|---|---|
| (Intercept: Baseline self-efficacy for restudy) | 852.87* | 40.47 |
| Time | 183.05* | 30.07 |
| Group: Retrieval Practice | 33.03 | 57.24 |
| Time × Group | −28.10 | 42.53 |
| Random effects | Variance | Std. Dev |
| Random slope of time per subject | 15,444.00 | 124.30 |
| Residual | 20,314.00 | 142.50 |
| Goodness of fit statistics | | |
| *AIC* (Akaike information criterion) | 3,370.30 | |
| *BIC* (Bayesian information criterion) | 3,391.30 | |
| Deviance | 3,358.30 | |

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < 0.05$, $^{\dagger}p < .1$

Self-efficacy ratings did not differ between groups at baseline (restudy: $M = 901.26$, $SD = 145.04$, retrieval practice: $M = 904.65$, $SD = 144.74$), $t = 0.58$, which suggests that the groups were equivalent on this measure at the beginning of the experiment. Self-efficacy ratings significantly increased over time, $t = 6.09$, but the rate of change did not differ significantly between retrieval practice group (increase of 154.95) and restudy group (increase of 183.05), as indicated by the nonsignificant interaction term, $t = 0.07$ (see Table 3). These results support our first hypothesis that self-efficacy would increase over time; however, given that the rate of change was similar for both the restudy and retrieval practice groups, our second hypothesis (that self-efficacy would increase more for the retrieval practice group) was not supported.

**Table 4** Summary of regression model assessing overall final test performance for Experiment 1

| | Estimate | Std. Error |
|---|---|---|
| (Intercept: Final test performance for restudy) | 75.01*** | 1.77 |
| Group: Retrieval practice | 6.77** | 2.50 |
| Change in self-efficacy | 4.42* | 1.81 |
| Group × Change in Self-Efficacy | −1.20 | 2.52 |
| $R^2$ | 0.12 | |
| Adjusted $R^2$ | 0.10 | |
| Num. obs. | 124 | |
| *RMSE* (root mean square error) | 13.92 | |

***$p < .001$, **$p < .01$, *$p < .05$, †$p < .1$

change in self-efficacy did not differ between restudy and retrieval practice groups.[3] These results indicate that there was an overall retrieval practice effect and that change in self-efficacy was positively related to memory performance.

To further assess hypothesis three (that we would see a standard retrieval practice effect), we ran two additional regression models to examine final test performance for previously seen items and new items, respectively. We were particularly interested in whether the retrieval practice effect would extend to new items, given that past work has shown mixed findings on whether performance on new items improves under retrieval practice conditions (see Adesope et al., 2017; Pan & Rickard, 2018; Rowland, 2014, for reviews). In the first model, performance on previously seen items was analyzed with group as a predictor and change in

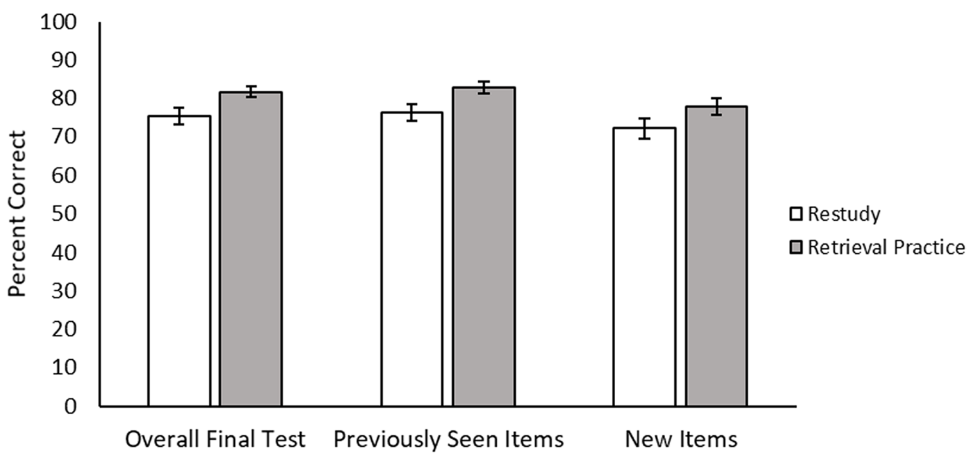


**Fig. 3** Memory performance as a function of group (restudy, retrieval practice) and item type (overall final test, previously seen items, and new items) for Experiment 1. Error bars represent standard error of the mean. The retrieval practice group performed significantly better than the restudy group on the overall final test and on previously seen items. The difference in memory performance between the retrieval practice group and restudy group for new items was marginally significant

To test our third hypothesis that there would be a standard retrieval practice effect, we performed a multiple regression analysis. For the regression model, final test score (all items) was the outcome variable, with group and change in self-efficacy as predictors (see Table 4 for a summary of the results). Overall, the regression model was significant, $F(3,120) = 5.28$, $p = .002$, $R^2 = .095$. Results showed that the retrieval practice group ($M = 81.60$, $SD = 11.67$) scored significantly higher on the final test compared with the restudy group ($M = 75.25$, $SD = 16.58$), $t = 2.70$, $p = .008$, which is in line with the typical retrieval practice effect (Adesope et al., 2017). Interestingly, change in self-efficacy was a significant positive predictor of final test score (across all participants), with larger increases (in self-efficacy) associated with higher final test scores, $b = 4.42$, $t = 2.45$, $p = .02$. The interaction between group and change in self-efficacy was not significant, $t = -0.48$, $p = .63$, suggesting that the degree of self-efficacy as a covariate since it was significantly related to overall final test performance (see Fig. 3 and Table 5 for a summary of results). This model was significant, $F(2, 121) = 6.45$, $p = .002$, $R^2 = .08$. Specifically, the retrieval practice group ($M = 82.86$, $SD = 12.21$) showed significantly higher memory than the restudy group ($M = 76.28$, $SD = 17.41$) as expected, $t = 2.63$, $p = .01$. Thus, change in self-efficacy was a significant predictor of participants' scores

[3] In our primary regression model, we used the self-efficacy change score as the dependent measure, but it is possible that baseline (T1) self-efficacy and final (T3) self-efficacy might be better predictors of memory performance. Thus, to ensure that we were using the self-efficacy measure that was the best predictor, we ran two additional models using T1 and T3 self-efficacy as predictors. Neither of these measures significantly predicted final test performance (*ps* > .05)

**Table 5** Summary of regression models assessing performance on previously seen versus new items for Experiment 1

| | Previously Seen Items | | New Items | |
|---|---|---|---|---|
| | Estimate | Std. Error | Estimate | Std. Error |
| (Intercept: Test performance for restudy) | 76.09*** | 1.87 | 71.91*** | 2.27 |
| Experimental group: Retrieval practice | 6.96** | 2.64 | 6.18† | 3.21 |
| Change in self-efficacy | 3.43* | 1.33 | 4.93** | 1.61 |
| $R^2$ | 0.10 | | 0.09 | |
| Adjusted $R^2$ | 0.08 | | 0.08 | |
| Num. obs. | 124 | | 124 | |
| *RMSE* (root mean square error) | 14.70 | | 17.85 | |

***$p < .001$, **$p < .01$, *$p < .05$, †$p < .1$

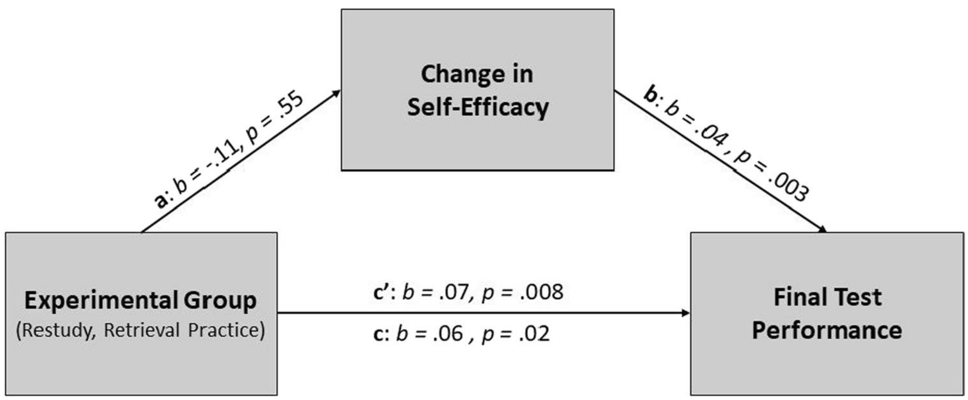


**Fig. 4** Pathway weights of the mediation model for Experiment 1. Path c is the direct effect, or the relationship between experimental group (restudy, retrieval practice) and memory on the final test not controlling for the mediator (self-efficacy). Path c' is the relationship between the experimental group and memory after controlling for the mediator. One can conclude mediation occurs when there is a significant drop from c to c'. Both experimental group (path c) and change in self-efficacy (path b) significantly predicted final test performance. Experimental group was not significantly related to change in self-efficacy (path a), however, and so the overall mediation was not significant

on previously seen items, $b = 3.43$, $t = 2.58$, $p = .01$, where larger increases in self-efficacy were related to higher final test performance. Next, in the second model, we analyzed performance for the new items. Results showed the overall model was significant, $F(2, 121) = 6.23$, $p = .003$, $R^2 = .08$. For new items, the retrieval practice group ($M = 77.82$, $SD = 16.07$) scored marginally higher than the restudy group ($M = 72.18$, $SD = 20.55$), $t = 1.93$, $p = .056$. The relationship between change in self-efficacy and test performance was also significant for new items, $b = 4.93$, $t = 3.06$, $p = .003$, with larger increases in self-efficacy being associated with better memory performance on the final test.

To test the fourth hypothesis about whether changes in self-efficacy partially accounted for the retrieval practice effect, we conducted a causal mediation analysis using the bias-corrected and accelerated (BCa) method. This method involves bootstrapping confidence intervals for the effects in the mediation analysis (direct, indirect, total effect) to determine whether a variable is a significant mediator. [4] In this analysis, treatment group was the predictor variable, final test performance was the outcome variable, and change in self-efficacy was the mediating variable. The bootstrapped confidence intervals for the mediation analysis were calculated based on 2,000 simulations. Results indicated that there was a significant direct effect of group on final test performance, $p = .01$, CI [0.02, 0.12] but no significant mediating effect of change in self-efficacy, $p = .53$, CI [−0.02, 0.01] (see Fig. 4 and Table 6 for more statistical information). This result shows that participants' degree of change in self-efficacy did not significantly contribute to the memory

[4] Work suggests that the BCa method of bootstrapping is better in terms of power and Type I error rate when used in small to moderate samples compared with other bootstrapping methods (Briggs, 2007; Preacher & Hayes, 2008).

**Table 6** Results of BCa mediation analysis for Experiment 1

| | β | 95% Confidence Interval |
|---|---|---|
| ACME (Average causal mediation effects) | −.004 | [−.02, .01] |
| ADE (Average direct effects) | .07* | [.02, .12] |
| Total effect | .06* | [.02, .12] |
| Proportion mediated | −0.07 | [−1.24, .09] |

2,000 simulations

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{\dagger}p < .1$

benefit from retrieval practice. Importantly, although not a significant mediator, change in self-efficacy was still a significant predictor of memory performance.

## Experiment 1 discussion

In Experiment 1, we collected preliminary evidence examining the relationship between retrieval practice, self-efficacy, and memory. We found that self-efficacy increased across the experiment at a similar rate for both retrieval practice and restudy groups. Retrieval practice alone may not have been a sufficient mastery experience to increase participants' self-efficacy above and beyond what they would normally experience during learning, at least under the conditions we used in this experiment. Further, we replicated the standard retrieval practice effect and found that these memory benefits may also extend beyond the items retrieved during learning to novel items. This finding points to the effectiveness of retrieval practice. Finally, although changes in self-efficacy did not explain the memory benefits due to retrieval practice, self-efficacy was positively related to memory performance as shown in prior work in this area (Beaudoin & Desrichard, 2011). Specifically, higher self-efficacy ratings were related to better memory on the final test. Because of the novelty of our research question (relationship between retrieval practice, self-efficacy, and memory), we conducted a second experiment to further investigate the relationship between retrieval practice, self-efficacy, and memory. In particular, we focused on an experimental approach to give participants false feedback on their performance on the retrieval practice task. We reasoned that providing feedback after engaging in retrieval practice would serve as a strong mastery experience signal which would increase the likelihood of seeing a relationship between retrieval practice, self-efficacy, and memory, if such a relationship exists. Although it is not common to give false feedback in retrieval practice investigations, using false negative (and positive) feedback is more common in experiments investigating the impact of self-efficacy on performance across different domains (Bandura & Jourden, 1991; Fitzsimmons et al., 1991; Litt 1988; Prussia & Kinicki, 1996). The logic behind this approach is that if there is truly a connection between self-efficacy and performance (of some type), then it should be the case that providing false feedback will affect self-efficacy and in turn modulate performance, delineating the connection between self-efficacy and behavior. Overall, Experiment 2 will allow us to better understand the connection between retrieval practice, self-efficacy, and memory.

# Experiment 2

In Experiment 2, we used an approach to experimental manipulate self-efficacy by giving false feedback during retrieval practice (negative, neutral, or positive feedback, regardless of actual performance in the retrieval practice tests) as a more rigorous procedure to understand the relationship between retrieval practice, self-efficacy, and memory. Our rationale for this approach (i.e., providing feedback) is that if retrieval practice does indeed have an influence on self-efficacy (and memory), then it should be the case that providing negative feedback (regardless of how participants actually perform during retrieval practice) should lead to reduced self-efficacy which then should yield poorer memory on a final test relative to a neutral feedback condition. We had three hypotheses for Experiment 2. First, we predicted that self-efficacy would change over time as we observed in Experiment 1. Second, we expected that the direction of change would differ based on the type of feedback participants received. Specifically, we predicted that self-efficacy would increase for the neutral and positive feedback group but decrease for the negative feedback group. Third, we hypothesized that memory performance would differ by feedback group such that changes in self-efficacy (due to the feedback) would partially account for memory on the final test. In particular, if negative feedback decreases self-efficacy, this should lead to a decrease in memory performance on a final test compared with the neutral feedback group. Further, if positive feedback increases self-efficacy significantly more than the neutral feedback group, we would expect corresponding improvements in memory above the neutral feedback group; however, if the neutral and positive groups show similar self-efficacy, we would not expect to see any differences in memory performance. Overall, Experiment 2 will provide a more rigorous test of the connection between retrieval practice, self-efficacy, and memory.

## Method

### Participants

One hundred and fifty-one participants from the university's psychology subject pool completed Experiment 2 (54% female, Age: $M = 19.50$, $SD = 1.88$). Approximately 40%

of participants identified as Asian or Pacific Islander, 21% as Latinx/Hispanic, 17% as White, 9% as Middle Eastern or North African, 4% as Black or African American, and 9% as other races or multiracial. Each participant was randomly assigned to a group (negative, neutral, or positive feedback).

### Materials

Experiment 2 used the same lecture videos, retrieval practice items, and final test items as Experiment 1.

**Feedback** We randomly assigned each participant to one of three feedback groups: negative feedback, neutral feedback, or positive feedback. Those in the negative feedback group received the message "Compared with others who have participated in the study, you have performed among the lowest." The neutral feedback group received the message "Compared with others who have participated in the study, you have performed about average." Finally, participants in the positive feedback group received the sentence "Compared with the others who have participated in the study, you have performed among the highest."

### Procedure

This experiment was a mixed design with feedback as the between-subjects factor and self-efficacy as a repeated-measures factor. There were two experimental feedback groups (negative and positive) and a control feedback group (neutral). Like Experiment 1, Experiment 2 was composed of two main phases: encoding and reexposure phase (Session 1) and a final test phase starting 48 hours later (Session 2). Unlike Experiment 1, Experiment 2 was conducted fully online and administered via Qualtrics survey (due to the COVID-19 pandemic).

**Encoding and reexposure phase** Experiment 2 had a similar structure to Experiment 1 but differed in two ways. First, participants were randomly assigned to one of the three feedback groups (negative, neutral, positive). Second, all participants engaged in a retrieval opportunity (no restudy group) after each lecture video. After each retrieval practice item, participants were shown the correct answer. Once all retrieval items were completed, participants received the feedback associated with their assigned group. The same feedback message was shown after each of the four encoding/reexposure cycles. At the end of Session 1, participants were reminded of Session 2 and were granted credit for their participation.

**Final test phase** Forty-eight hours after Session 1, participants completed Session 2 (i.e., final test). The same final test items from Experiment 1 were used in Experiment 2. Participants completed the final test items at their own pace, with one item presented on the screen at once. After completing the final test, participants answered demographic questions, received a debriefing document, and were granted credit for their participation.

**Data analysis** All analyses were conducted using the statistical computing software R. We used the *lme4* package (Bates et al., 2015) for the mixed-effects model, and *mediation* package (Tingley et al., 2014) for the mediation analysis.

## Results

We report self-efficacy as a function of time (T1, T2, T3) and feedback group (negative, neutral, positive) in Fig. 5. To test the first and second hypotheses (that self-efficacy would increase over time and that this increase would be different based on the type of feedback received), we conducted a mixed-effects regression. We tested a mixed-effects model with self-efficacy score as the dependent variable, and feedback group (between-subjects factor: negative, neutral, positive) and time (repeated-measure: T1, T2, and T3) as fixed predictors. The model included each subject's intercept (mean self-efficacy score) and slope of change over time as random factors to account for individual differences in mean self-efficacy and how participants' self-efficacy changed across Session 1. Since the neutral feedback group was designed to be a control, it was used as the referent group in this analysis. Results showed that self-efficacy significantly increased over time, $t$ = 7.05, and importantly, the slope of this change differed based on feedback group (see Table 7). Compared with the neutral feedback group, the negative feedback group had a significantly more negative slope, $t = -6.14$, whereas the slope for the positive feedback group did not significantly differ from the neutral feedback group, $t = -1.86$. These results support our first two hypotheses: that self-efficacy would change over time and that the direction of change would differ based on the feedback participants received.

To assess our third hypothesis about whether memory performance differed by feedback group, and further, whether changes in self-efficacy partially accounted for those differences, we conducted causal mediation analyses using the same bootstrapping method as in Experiment 1. Figure 6 shows memory performance as a function of feedback group. We first examined memory for the three different feedback conditions. Results showed that the negative feedback group had significantly worse memory performance than the neutral feedback group, $t = -2.46$, $p = .02$, but there was no significant difference in memory performance between the positive feedback group and neutral feedback group, $t = -1.25$, $p = .21$. Further, there was no significant difference in memory

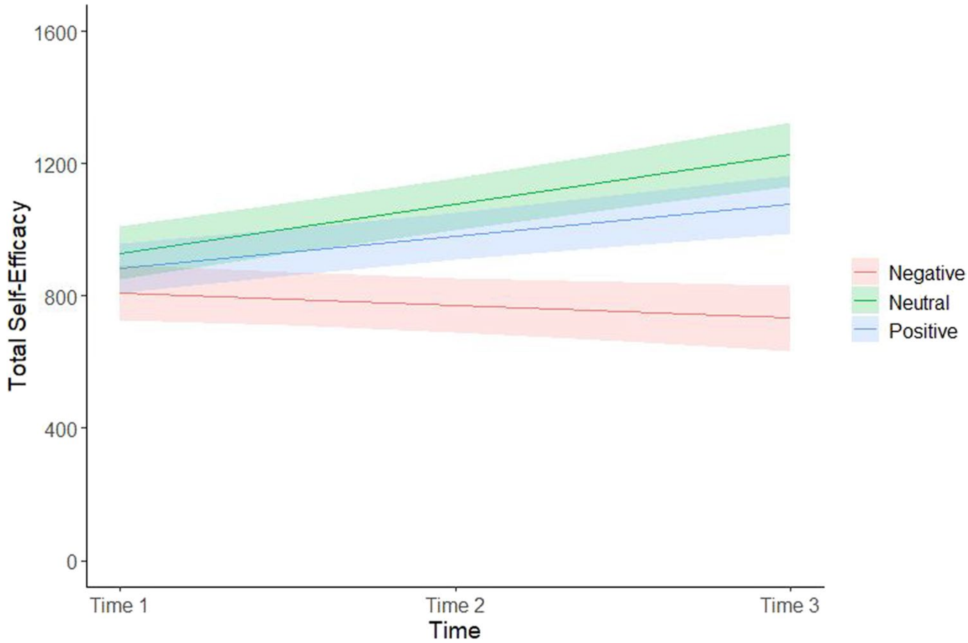


**Fig. 5** Change in self-efficacy by time point and feedback group for Experiment 2. Neutral and positive feedback groups showed increased self-efficacy over time and at a similar rate. The negative feedback group, however, showed significant decreases in self-efficacy over time

**Table 7** Summary of mixed-effects model assessing change in self-efficacy in Experiment 2

| Fixed Effects | Estimate | Std. Error | |
|---|---|---|---|
| (Intercept: Baseline self-efficacy for neutral) | 929.32*** | 40.94 | |
| Time | 148.92*** | 21.12 | |
| Feedback group: Negative | −119.41* | 59.17 | |
| Feedback group: Positive | −43.98 | 55.83 | |
| Time × Negative Group (slope) | −187.35*** | 30.53 | |
| Time × Positive Group (slope) | −53.51 | 28.81 | |
| Random effects | Variance | Std. Dev. | Correlation |
| Random Intercept per subject | 73,977.53 | 271.99 | |
| Random slope of time per subject | 16,978.55 | 130.30 | −.24 |
| Residual | 9,766.67 | 98.83 | |
| Goodness of fit statistics | | | |
| *AIC* (Akaike information criterion) | 6165.30 | | |
| *BIC* (Bayesian information criterion) | 6206.50 | | |
| Deviance | 6145.30 | | |

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < 0.05$, $^{\dagger}p < .1$

performance between the negative and positive feedback groups, $t = 1.32$, $p = .19$.[5]

Because there were three feedback groups, two mediation analyses were required to compare each treatment group (negative and positive) to the neutral feedback group (control) to examine the extent that self-efficacy partially accounted for memory differences. In both analyses, the predictor variable was treatment group, final test performance was the outcome variable, and change in self-efficacy was the mediating variable. The first mediation analysis contrasted the negative feedback group with

[5] If self-efficacy impacts memory performance through persistence, then it is possible that response times might differ across the three different feedback groups. Specifically, it is possible that those in the neutral and positive feedback groups would spend longer on the final test compared with those in the negative feedback group. Numeri-

Footnote 5 (continued)

cally, the positive feedback group showed the longest response times ($M$ = 2,531 ms), followed by the negative group ($M$ = 2,204 ms), and then the neutral group ($M$ = 1899 ms); however, these differences were not statistically significant, $F(2, 94) = 0.79$, $p = .46$.

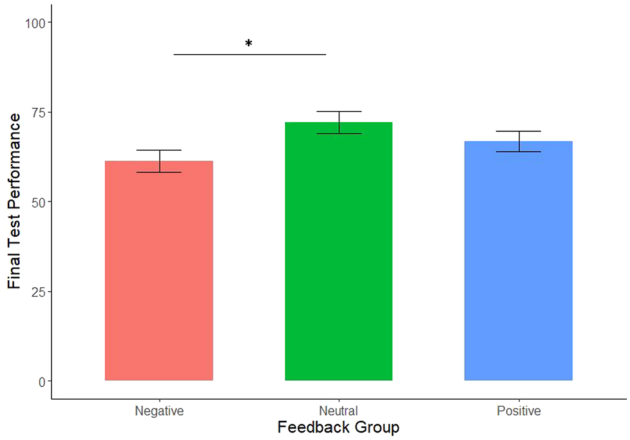


**Fig. 6** Memory performance on final test by feedback group for Experiment 2. The negative feedback group had significantly lower memory performance compared with the neutral feedback group. There were no other differences between groups

that after accounting for change in self-efficacy, there was not a significant direct effect of feedback group on final test performance, $p = .60$, CI [-6.76, 10.49]. This finding, coupled with a significant indirect effect (mediating effect), $p < .001$, CI[−16.59, −6.24], indicates that changes in self-efficacy explain a significant amount of variance in final test performance between the negative and neutral feedback groups. Specifically, decreases in self-efficacy for the negative feedback group led to worse memory performance compared with the neutral feedback group.

The second mediation analysis contrasted the positive feedback group with the neutral feedback group. Because we did not see greater self-efficacy for the positive compared with neutral feedback conditions, we did not expect that self-efficacy would mediate the relationship between group and memory. Figure 8 and Table 8 contain the statistical information for this analysis. The total effect of the model was not significant, $p = .32$, CI [−12.15, 3.54], and there was not a significant direct effect of feedback

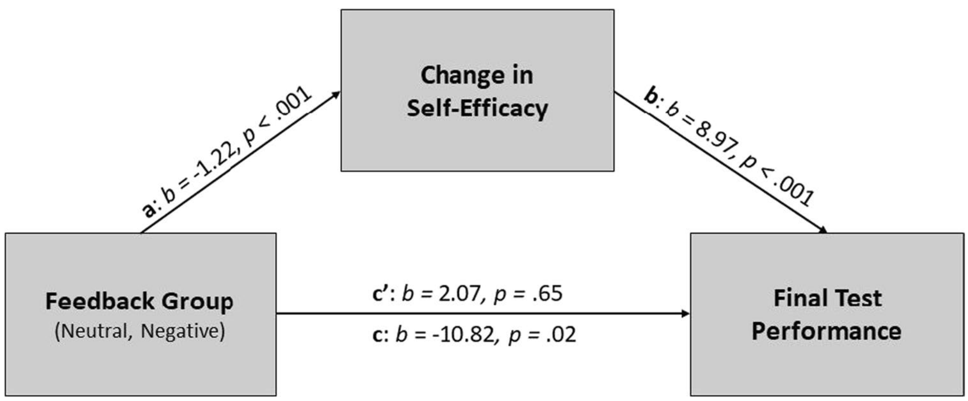


**Fig. 7** Pathway weights of the first mediation model in Experiment 2 (negative feedback compared with neutral feedback). Feedback group significantly predicted change in self-efficacy (path a). Change in self-efficacy (path b) and feedback group (path c) significantly predicted final test performance. The change in the direct effect (c') after accounting for change in self-efficacy indicates that there is a significant mediating effect of change in self-efficacy

**Table 8** Results of BCa mediation analyses for Experiment 2

| | Neutral-Negative | | Neutral-Positive | |
|---|---|---|---|---|
| | β | 95% CI | β | 95% CI |
| ACME (Average causal mediation effects) | −10.98*** | [−16.59, −6.24] | −3.35* | [−6.65, −.83] |
| ADE (Average direct effects) | 2.07 | [−6.76, 10.49] | −.84 | [−8.73, 6.21] |
| Total effect | −8.90* | [−17.18, −1.24] | −4.19 | [−12.15, 3.54] |
| Proportion mediated | 1.23* | [.91, 627.99] | .80 | [−202.09, .32] |

2,000 simulations

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{\dagger}p < .1$

the neutral feedback group. Fig. 7 and Table 8 contain the statistical information for this analysis. The total effect of the model was significant, $p = .03$, CI [−17.18, −1.24]. Results indicated group on final test performance, $p = .85$, CI [−8.73, 6.21]. This indicates that there was no relationship

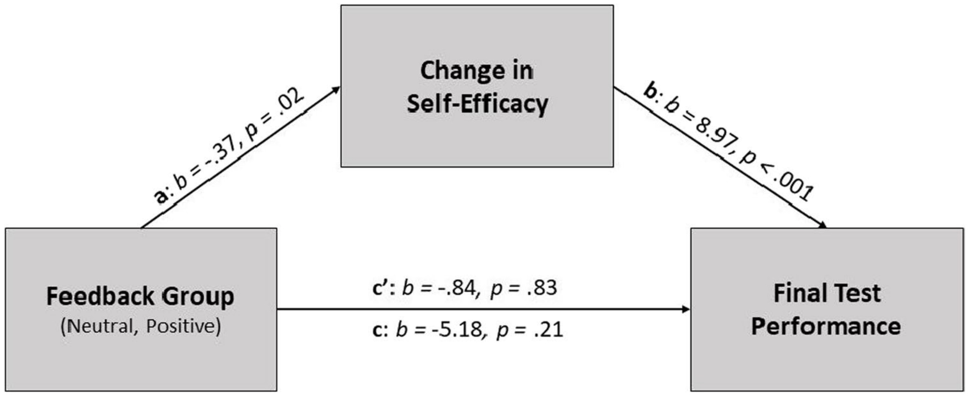


**Fig. 8** Pathway weights of the second mediation model in Experiment 2 (positive feedback compared with neutral feedback). Feedback group significantly predicted change in self-efficacy (path a) and change in self-efficacy (path b) significantly predicted final test performance. However, condition was not significantly related to final test performance (path c), and so the overall mediation was not significant

between feedback group and memory performance, and thus no mediating effect of change in self-efficacy.

The two mediation analyses taken together provide a measure of support for our third hypothesis: that for participants in the negative feedback group, decreases in self-efficacy accounted for lower memory performance compared with the neutral feedback group. Importantly, such a finding suggests a connection between change in self-efficacy due to engaging in retrieval practice and memory performance on a final test, especially when given negative feedback.

### Experiment 2 discussion

Overall, Experiment 2 provides evidence of a relationship between retrieval practice, self-efficacy, and memory, especially in the case where participants are given false negative feedback. Specifically, our findings suggest that negative feedback led to significantly worse memory than the neutral feedback group which was partially explained by decreased self-efficacy. There were no differences between the neutral and positive feedback groups in terms of change in self-efficacy or memory performance. Even with all participants engaging in an effective memory strategy (retrieval practice), decreases in self-efficacy still impacted how well participants performed on the final test, which offers evidence in favor of a relationship between retrieval practice, self-efficacy, and memory.

## General discussion

In these experiments, we investigated the extent self-efficacy contributes to the memory benefit from retrieval practice. There were four main findings. First, across all experiments self-efficacy increased significantly over time (with the exception of the negative feedback group in Experiment 2). Second, in Experiment 1, we found a significant retrieval practice effect where final test performance was better in the retrieval practice group compared with restudy group. This is consistent with past work demonstrating that retrieval practice effectively improves memory over restudy (e.g., Adesope et al., 2017; Pan & Rickard, 2018; Rowland, 2014). Third, and perhaps most importantly, we found that self-efficacy mediated the relationship between self-efficacy and memory in the negative feedback group compared with the neutral feedback group in Experiment 2. Such a finding suggests a potentially underrecognized relationship between retrieval practice, self-efficacy, memory. Interestingly, self-efficacy did not mediate the relationship between retrieval practice and memory on the final text in Experiment 1, which may suggest that retrieval practice without additional feedback may not be a sufficiently strong mastery experience to affect self-efficacy (and subsequently memory), at least under the experimental conditions we used in Experiment 1. Fourth, in both experiments we found evidence that self-efficacy significantly predicted memory on the final test, where increases in self-efficacy were related to increases in memory. Such a finding may represent a promising mechanism to induce memory improvement in people through modulating self-efficacy. Taken together, this set of experiments provides evidence of a previously unrecognized relationship between retrieval practice, self-efficacy, and memory, which is a promising area for future research.

Starting first with Experiment 1, one of our primary aims was to investigate whether self-efficacy would improve over time and whether this change in self-efficacy would be larger for those in the retrieval practice group compared with restudy. Prior work has shown that interventions during learning can improve self-efficacy (see Huang & Mayer, 2018), and one effective method to increase self-efficacy is by engaging in mastery experiences (Bandura, 1977). We hypothesized that retrieval practice could serve as a mastery experience, which would lead to larger gains in self-efficacy compared with the restudy group. Although self-efficacy did increase over time, this change in self-efficacy was of similar magnitude in both the restudy and retrieval practice groups, in contrast to our predictions. We see two potential explanations why self-efficacy levels were similar for both groups. First, one possibility is that because participants in the control group did not engage in retrieval practice, their self-efficacy ratings may have reflected overconfidence in their ability to remember information at final test (see Miller & Geraci, 2014). If it is the case that participants in the restudy group had inflated self-efficacy ratings, consistent with prior work, then this artificial increase in self-efficacy may have obscured our ability to detect a true increase in self efficacy due to retrieval practice effect. Second, a different possibility is that retrieval practice may not have acted as a sufficiently strong mastery experience to modulate self-efficacy, at least under the retrieval practice conditions we used in Experiment 1. Although retrieval practice attempts simulated the final test, participants may not have received enough diagnostic information about their performance (i.e., feedback) to truly make retrieval practice a sufficient mastery experience to boost self-efficacy in Experiment 1. Importantly, in Experiment 2, we experimentally manipulated self-efficacy by having all participants engage in retrieval practice and providing either negative, neutral, or positive feedback (regardless of performance) on the retrieval practice task. Critically, results of Experiment 2 showed that negative feedback reduced self-efficacy, and that this reduction in self-efficacy partially accounted for reduced memory on the final test. Given that past work has experimentally manipulated self-efficacy as a means to find stronger causal links between self-efficacy and performance in different domains (Bandura & Jourden, 1991; Fitzsimmons et al., 1991; Litt, 1988; Prussia & Kinicki, 1996), the results of Experiment 2 provide evidence of a relationship between retrieval practice, self-efficacy, and memory, especially under negative feedback conditions. Such a finding is consistent with the idea that retrieval practice, especially retrieval practice with feedback, serves as a mastery experience that has a strong influence on self-efficacy, and that these changes in self-efficacy have a direct effect on memory performance. It is worth noting that the design we used in Experiment 2 eliminated the potential issue of overconfidence in a restudy group (which could have obscured our ability to better understand the relationship between self-efficacy and memory in Experiment 1). Interestingly, memory performance did not differ between the positive and neutral feedback conditions in Experiment 2. One possible reason for this may be related to past findings in self-efficacy studies that try to boost self-efficacy. This work has shown a paradoxical effect where conditions meant to increase self-efficacy can lead to overall poorer performance on a task of interest (Bandura & Locke, 2003). The rationale behind such so-called negative self-efficacy effects is that conditions that are meant to strongly increase self-efficacy, can sometimes lead individuals to underperform on a task (which can happen for a variety of reasons such as poor execution of the task, or reduced motivation to complete the task; Bandura & Locke, 2003). Overall, although Experiment 2 showed evidence of the relationship between retrieval practice, self-efficacy, and memory, future work should focus on manipulating feedback for both retrieval practice and restudy groups to better understand the effects on self-efficacy on memory under retrieval practice experimental conditions. Overall, finding a relationship between retrieval practice, self-efficacy, and memory, adds to other memory-related work that has the potential to improve memory in educational contexts (Butler & Rodiger, 2007; Giannakopoulos et al., 2021; McCurdy et al., 2021; McCurdy et al., 2019; McCurdy et al., in press; McCurdy, Viechtbauer, et al., 2020; McDaniel et al., 2007). One practical implication of this work is that conditions that give (or imply) negative feedback to students in educational settings may not be motivating, but instead may have the unintended effect of leading to poorer performance on academic tasks.

In addition to examining the relationship between self-efficacy and retrieval practice, we were also interested in examining the extent self-efficacy would be positively related to memory performance overall. Previous work has shown that higher self-efficacy is associated with better memory for word lists (Beaudoin & Desrichard, 2011) and better academic performance (Multon et al., 1991; Robbins et al., 2004). In line with this limited literature, results from Experiment 1 showed that self-efficacy positively predicted memory performance. Specifically, people who had a greater degree of change in self-efficacy also had higher scores on the final test. Further, in Experiment 2, we found a similar relationship where increases in self-efficacy were related to better memory whereas decreases in self-efficacy were related to poorer memory. It is worth noting that we measured self-efficacy by asking participants to rate their confidence (0-100) in their ability to complete various tasks related to research methods. Because of this approach it could be argued that self-efficacy as measured in this study might be related to a measure in the metacognition literature: judgments of learning (JOLs). In JOL experiments, participants study lists of words or other materials and are subsequently asked (either immediately or after a delay) to rate how likely they would be to remember *that* stimuli, often on a 0-100% scale (e.g., Arbuckle & Cuddy, 1969; Rhodes & Castel, 2008). Accuracy of JOLs (how well participants can predict whether they will later remember an item) is then measured by correlating participants' original JOLs with their performance on a final memory test. Although it is possible there is overlap between JOLs and self-efficacy because of the similar rating scale, we argue that self-efficacy substantively differs from JOLs because self-efficacy is not focused on predicting accuracy on future memory tasks, per se. Rather, self-efficacy is an individual's *confidence* in their ability to do well on specific tasks, which we think is an important theoretical distinction. Based on this line of evidence, we view JOLs and self-efficacy are separate constructs. Future work should directly test the relationship of these constructs (self-efficacy and JOLs) to memory performance to provide stronger evidence that these are distinct psychological phenomena.

As a part of Experiment 1, we intended to replicate previous work showing the benefits of retrieval practice on memory (e.g., Adesope et al., 2017; Pan & Rickard, 2018; Rowland, 2014). In addition to a standard retrieval practice effect in Experiment 1 (e.g., better memory for previously tested items compared with restudied items), we found improved memory for new items that participants did not see during encoding, though this effect was less robust (marginal). This is consistent with work showing that the retrieval practice effect is robust for previously seen items, but more mixed for new items (e.g., Adesope et al., 2017; Pan & Rickard, 2018). The fact that we saw a marginally significant effect for new items suggests that there may have been transfer of learning to previously untested material due to retrieval practice. This is in line with the elaborative retrieval hypothesis (Carpenter, 2009), which states that retrieving information from memory strengthens semantically related information, making it easier to retrieve at final test. Overall, attempting to engage in retrieval practice may have improved memory for conceptually related materials (i.e., the new items on the final test). Future work should focus on better understanding how and when the benefit of retrieval practice extends to previously untested items.

Though cognitive mechanisms have been proposed to explain the retrieval practice effect, we were interested in the extent self-efficacy, a social-cognitive factor, could contribute to this memory benefit. It is well known constructs related to one's sense of self have a strong effect on memory (Brown et al., 1986; Gutchess et al., 2007; Ilenikhena et al., 2021; Leshikar et al., 2015b; Leshikar & Duarte, 2012, 2014; Rogers et al., 1977; Symons & Johnson, 1997; Wong et al., 2017), and past work has demonstrated a relationship between self-efficacy and memory in both laboratory and academic contexts (Beaudoin & Desrichard, 2011; Huang & Mayer, 2018; Multon et al., 1991; Robbins et al., 2004). The work we present in this investigation provides additional evidence that changes in self-efficacy is related to memory performance. In Experiment 2, we directly manipulated feedback during retrieval practice to better understand whether self-efficacy directly leads to changes in memory performance on a final test. Importantly, we found that decreases in self-efficacy accounted for worse memory on the final test compared with the neutral feedback group, which implies that self-efficacy has an influence on memory, at least under negative feedback conditions. Prior research has found that higher levels of self-efficacy are associated with increased persistence and academic achievement (Multon et al., 1991; Robbins et al., 2004), as well as with better memory performance in older adults (Beaudoin & Desrichard, 2011). If increased self-efficacy results in greater task persistence, it is possible that this persistence may result in more extensive searches through memory during retrieval practice and final test. This in turn could lead to enhanced memory. Similarly, it may be that *reductions* in self-efficacy lead to less persistence in memory tasks, such as reduced search through memory, which in turn lead to poorer memory. Although speculative, it may be that getting feedback on their performance, especially negative feedback, leads participants to adjust their behavior during learning, and on the final test. For instance, getting negative feedback might make a participant aware that they are not doing well on the task at hand (i.e., completing the memory test), which may decrease self-efficacy. Decreased self-efficacy, in turn, may then lead to reduced processing of studied materials and/or reduced effort in practicing retrieval, which would result in poorer memory. Future work should investigate the specific mechanisms by which changes in self-efficacy modulates memory, including whether there is a link between self-efficacy and metacognition. Understanding how memory may be improved is an important pursuit (Bjork & Bjork, 2020; Burden et al., 2021; Frankenstein et al., 2020; Kadwe et al., 2022; Leach et al., 2018; Leshikar et al., 2016; Leshikar et al., 2012; Leshikar et al., 2015a; Leshikar & Gutchess, 2015; Leshikar et al., 2017; Matzen et al., 2015; Meyer et al., 2020; Patel et al., in press; Sklenar et al., 2021; Udeogu et al., in press; Villasenor et al., 2021), and the current study contributes to this body of knowledge.

Although we found evidence that changes in self-efficacy directly affect final memory performance, there are six limitations of our experiments worth describing. First, in both experiments, we investigated the relationship between retrieval practice, self-efficacy, and memory with educational materials that were similar to what participants might encounter in the classroom. With this decision, we introduced the possibility that participants may have had preexisting conceptions and self-efficacy related to these topics that could have potentially clouded the relationship between our constructs of interest. We see this as less likely given that participants in these experiments were taking an introductory psychology class and thus would not likely have extensive exposure to research methods content. Future work should investigate the relationship between retrieval practice, self-efficacy, and memory using different content to extend the findings of this report. Second, another potential limitation was our use of multiple-choice items during encoding and reexposure phase for both the retrieval practice and restudy groups in Experiment 1. Multiple-choice formatted information is not typically used for the restudy group in retrieval practice paradigms, and this could have allowed participants to engage in covert retrieval which could have affected their self-efficacy ratings (i.e., they were less overconfident than typical restudy groups), which could have reduced our ability to truly measure self-efficacy differences between the retrieval practice and restudy groups. Although our aim with this procedure (using multiple choice items) was to maintain uniformity in the material seen by both the retrieval practice and restudy groups, it is possible this procedure may have affected our findings. Importantly, we were able to address this potential issue in Experiment 2 by making retrieval practice a part of the procedure for all participants and experimentally manipulating feedback instead. Future work can examine whether feedback or other messages are effective means of reducing potential overconfidence in both retrieval practice and restudy groups, respectively. Third, in both experiments, we used an interpolated retrieval practice paradigm and assessed self-efficacy before, during, and after learning. Interspersing retrieval opportunities and self-efficacy ratings throughout learning may have allowed participants to better gauge their own self-efficacy compared with a standard testing paradigm where participants are only tested once after encoding. Further, by asking participants to rate their self-efficacy at baseline on topics related to studied materials, this may have led participants to pay special attention to those topics while watching the video lessons. Future research should explicitly look at whether interpolated retrieval practice is more or less effective at increasing self-efficacy compared with single retrieval paradigms (i.e., not interpolated) where self-efficacy is not measured before learning. Fourth, in Experiment 2, we did not include a restudy condition. Because of this experimental approach, it is not possible to make claims on whether retrieval practice effects in memory can be eliminated (or reversed) with sufficiently poor self-efficacy. Future work could use a similar procedure to Experiment 2, but include an additional restudy control condition, to examine the extent the retrieval practice effect can be eliminated or reversed entirely. Fifth, although we found evidence of the relationship between retrieval practice, self-efficacy, and memory (particularly Experiment 2), it is possible to argue that self-efficacy may have a limited effect on memory benefits from retrieval practice, which would fit the findings of Experiment 1 (i.e., that self-efficacy did not mediate the relationship between condition and memory) as well as our findings that there was not difference in memory between the positive and neutral feedback conditions in Experiment 2. Although this alternative explanation is plausible, we think the approach we used in Experiment 2 of providing false feedback follows a gold standard approach in the self-efficacy literature to strongly link self-efficacy with performance across different domains. Sixth, we gave false feedback in Experiment 2. One downside to this approach is that participants may have been surprised to get feedback out of line with their performance, which may have affected results. Future work should further investigate the relationship between self-efficacy, retrieval practice, and memory using veridical feedback.

Results of this set of experiments shows a previously unrecognized relationship between retrieval practice, self-efficacy, and memory. As prior work has shown, retrieval practice can be easily incorporated in applied settings (Bangert-Drowns et al., 1991; Leeming, 2002; Screbo et al., 1992; Szpunar et al., 2013), and the findings of this set of experiments suggest that interventions aimed at modulating self-efficacy through retrieval practice paradigms are a promising way to affect memory and learning outcomes. It is especially valuable given that the benefits of retrieval practice extended to previously untested items, which has important implications for improving student learning. Overall, findings of these experiments give a richer understanding of the relationship between retrieval practice, self-efficacy, and memory, which is a potentially rich area for future work. These findings, and related future work, have the potential to inform best practices in educational contexts.

## Appendix A

(Definitional)

What term refers to making an abstract concept more concrete so that it can be measured?

A Independent variable
B Dependent variable
C **Operationalizing**
D Manipulating

(Application)

Which of the following is a way to manipulate fatigue (tiredness)?

A Keep half the participants up all night and let the other half sleep normally
B Ask participants how many hours they slept the night before
C Ask half the participants to exercise for 30 minutes and ask the other half sit quietly for 30 minutes
D All of the above
E **A & C**

(Definitional)

Reliability is...

A Whether a measure is measuring what it's supposed to
B **Whether a measure is consistent**
C Whether a measure is abstract enough
D Whether a measure is concrete enough

(Application)
A researcher created a new measure and wants to make sure it's reliable. To do this, she asks a group of people to complete her measure one day, then come back a week later to complete the same measure again. What type of reliability is this?

A Split-half reliability
B **Test–retest reliability**
C Reversed forms reliability
D Alternate forms reliability

(Application)
A researcher is interested in learning about how university students in the library will react to disturbing noise. The researcher sits in a crowded area of the library and starts playing music from her laptop speakers. She then records students' reactions to the music. What type of research design is this?

A Experiment
B Observation
C **Participant observation**
D Correlation

(Application)
A researcher collected data on students' sleep schedules and their exam grades. Data showed that the more students slept, the better their exam scores. Which type of correlation does this illustrate?

A **Positive**
B Negative
C Zero correlation

(Definitional)
What is the main advantage of experimental research?

A Experiments in the lab are very similar to the real world
B **Experiments can show cause and effect relationships**
C Experiments can be used to study all research questions
D Experiments do not require a lot of resources

(Definitional)
Why is random assignment important for determining cause and effect?

A It makes sure people do not have an equal chance of being chosen
B It adds confounding variables to the experiment
C **It eliminates alternative reasons for changes in the dependent variable**
D It influences the participants' responses

## Appendix B

The following questions assess your confidence in your ability to complete certain tasks related to research methods. Please answer these questions based on your **current** levels of confidence, rather than how confident you think you might be later.

*Rate your degree of confidence by recording a number from 0 to 100 using the scale given below:*

| 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|
| Cannot do at all | | | | | Moderately certain can do | | | | | Highly certain can do |

| | **Confidence (0–100)** |
|---|---|
| If you were given an example research study, how confident are you that you would know what research design was being used? | |
| If you were given an example research study, how confident are you in your ability to identify at least one strength and one weakness of the study? | |
| If you were asked to point out the main characteristics of a research study, how confident are you that you would be able to do so? | |
| If you were told about a potential bias in a research study, how confident are you that you could come up with a solution to fix the bias? | |
| How confident are you in your ability to explain when researchers can make cause-and-effect claims? | |
| How confident are you in your ability to explain the importance of random assignment in a research study? | |
| If you were given an example research study, how confident are you that you could identify at least one potential confound in that study? | |
| If you were given an example research study, how confident are you that you could identify the independent and dependent variables in that study? | |
| If you were given a psychological construct (ex: happiness, anxiety), how confident are you in your ability to come up with at least one way to measure that construct? | |
| If you were given an example study, how confident are you that you could determine whether a variable is being measured or manipulated? | |

If you were given a psychological measure, how confident are you in your ability to come up with at least one way to test the reliability of the measure?

If you were given a psychological measure, how confident are you in your ability to tell whether the measure is valid?

Given an example correlation coefficient (ex: $r = 0.5$, $r = -0.3$), how confident are you in your ability to explain what the correlation coefficient means?

How confident are you that you can explain at least one limitation of correlational research?

Given an example of an observational study, how confident are you that you could determine what type of observation was being used?

How confident are you in your ability come up with at least one strength and one weakness of survey research?

## Appendix C

Four distractor tasks were used in this experiment: forward digit span, backward digit span, digit-symbol substitution (WAIS-R; Wechsler, 1981), and phonemic F-A-S fluency task (Spreen & Benton, 1977). The following is a more detailed description of each.

The forward digit span task involved participants listening to the experimenter read out series of numbers that increased in length. After each string of numbers, participants wrote down the numbers in the same order they heard them. Backward digit span involved a similar procedure: participants heard strings of numbers of increasing length, but their task was to write down each series of numbers in reverse order that they heard them. The forward and backward digit spans were scored together as the total number of digit strings that were written down correctly. In the digit-symbol substitution task (WAIS-R), participants saw digits from 1 to 9 that were paired with different symbols. Below were a series of boxes with digits above and blank boxes below. Participants had 90 seconds to fill in as many blank boxes as they could with the appropriate symbols. This task was scored as the total number of boxes that were filled in correctly. For the F-A-S fluency task, participants heard a letter of the alphabet and had 1 minute to write down as many words as they could that started with that letter. They completed this task for three letters: F, A, S. The number of valid words listed for each letter were totaled, excluding proper nouns, proper adjectives, or repeated words.

## References


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.

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology, 81*(1), 126–131.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Bandura, A. (1995). Exercise of personal and collective efficacy in changing societies. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 1–45). Cambridge University Press. https://doi.org/10.1017/CBO9780511527692.003

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-efficacy beliefs of adolescents, 5*, 307–337.

Bandura, A., & Jourden, F. J. (1991). Self-regulatory mechanisms governing the impact of social comparison on complex decision making. *Journal of Personality and Social Psychology, 60*(6), 941.

Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology, 88*(1), 87.

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. L. C. (1991). Effects of frequent classroom testing. *Journal of Educational Research, 85*, 89–99.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin, 137*(2), 211–241. https://doi.org/10.1037/a0022106

Betz, N. E., & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational Behavior, 23*, 329–345.

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition, 9*(4), 475.

Brown, P., Keenan, J. M., & Potts, G. R. (1986). The self-reference effect with imagery encoding. *Journal of Personality and Social Psychology, 51*(5), 897.

Briggs, N. E. (2007). Estimation of the standard error and confidence interval of the indirect effect in multiple mediator models. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 67*(8-B), 4755.

Burden, C., Leach, R. C., Sklenar, A. M., Urban Levy, P., & Frankenstein, A. N., & Leshikar, E.D. (2021). Examining the influence of brain stimulation to the medial prefrontal cortex on the self-reference effect in memory. *Brain and Behavior,* Article e2368. https://doi.org/10.1002/brb3.2368

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1118–1133. https://doi.org/10.1037/a0019902

Butler, A. C., & Roediger III, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*(4/5), 514–527.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1563–1569. https://doi.org/10.1037/a0017021

Delaney, P. F., Verkoeijen, P. P., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of Learning and Motivation, 53*, 64–111.

DeVellis, R.F. (2012). *Scale development: Theory and applications*. SAGE Publications.

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

Fitzsimmons, P. A., Landers, D. M., Thomas, J. R., & Van Der Mars, H. (1991). Does self-efficacy predict performance in experienced weightlifters? *Research Quarterly for Exercise and Sport, 62*(4), 424–431.

Frankenstein, A. N., McCurdy, M. P., Sklenar, A. M., Pandya, R., Szpunar, K. K., & Leshikar, E. D. (2020). Future thinking about social targets: The influence of prediction outcome on memory. *Cognition, 204*, 1–9. https://doi.org/10.1016/j.cognition.2020.104390

Giannakopoulos, K., McCurdy, M. P., Sklenar, A. M., Frankenstein, A. N., Urban Levy, P., & Leshikar, E. D. (2021). Less constrained practice tests enhance the testing effect for item memory but not context memory. *American Journal of Psychology, 134*(3), 321–332. https://doi.org/10.5406/amerjpsyc.134.3.0321

Gutchess, A. H., Kensinger, E. A., Yoon, C., & Schacter, D. L. (2007). Ageing and the self-reference effect in memory. *Memory, 15*(8), 822–837.

Huang, X., & Mayer, R. E. (2018). Adding self-efficacy features to an online statistics lesson. *Journal of Educational Computing Research, 57*(4), 1003–1037.

Ilenikhena, G., Narmawala, H., Sklenar, A. M., McCurdy, M. P., Gutchess, A. H., & Leshikar, E. D. (2021). STOP SHOUTING AT ME: The influence of case and self-referencing on explicit and implicit Memory. *Frontiers in Psychology, 12*, 685756. https://doi.org/10.3389/fpsyg.2021.685756

Kadwe, P. P., Sklenar, A. M., Frankenstein, A. N., Urban Levy, P., & Leshikar, E. D. (2022). The influence of memory on approach and avoidance decisions: Investigating the role of episodic memory in social decision making. *Cognition, 225,* Article 105072. https://doi.org/10.1016/j.cognition.2022.105072

Leach, R. C., McCurdy, M. P., Trumbo, M., Matzen, L. E., & Leshikar, E. D. (2018). Differential age effects of transcranial stimulation on associative memory in younger and older adults. *The Journals of Gerontology: Series B*. https://doi.org/10.1093/geronb/gby003

Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology, 29*, 210–212.

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 464–470.

Leshikar, E. D., Cassidy, B. S., & Gutchess, A. H. (2016). Similarity to the self influences cortical recruitment during impression formation. *Cognitive, Affective, & Behavioral Neuroscience, 16*(2), 302–314. https://doi.org/10.3758/s13415-015-0390-3

Leshikar, E. D., & Duarte, A. (2012). Medial prefrontal cortex supports source memory accuracy for self-referenced items. *Social Neuroscience, 7*(2), 126–145. https://doi.org/10.1080/17470919.2011.585242

Leshikar, E. D., & Duarte, A. (2014). Medial prefrontal cortex supports source memory for self-referenced materials in young and older adults. *Cognitive, Affective, & Behavioral Neuroscience, 14*(1), 236–252. https://doi.org/10.3389/fnhum.2013.00537

Leshikar, E. D., Duarte, A., & Hertzog, C. (2012). Task-selective memory effects for successful implemented encoding strategies. *PLOS ONE, 7*(5), e38160. https://doi.org/10.1371/journal.pone.0038160

Leshikar, E. D., Dulas, M. R., & Duarte, A. (2015a). Self-referencing enhances recollection in both young and older adults. *Aging, Neuropsychology, and Cognition, 22*(4), 388–412. https://doi.org/10.1080/13825585.2014.957150

Leshikar, E. D., & Gutchess, A. H. (2015). Similarity to the self affects impression memory for others. *Journal of Applied Research in Memory and Cognition, 4*(1), 20–28. https://doi.org/10.1016/j.jarmac.2014.10.002

Leshikar, E. D., Leach, R. C., McCurdy, M. P., Trumbo, M., Sklenar, A. M., Frankenstein, A. N., & Matzen, L. E. (2017). Transcranial direct current stimulation of dorsolateral prefrontal cortex improves recall but not recognition memory. *Neuropsychologia, 106*, 390–397.

Leshikar, E. D., Park, J. M., & Gutchess, A. H. (2015b). Similarity to the self affects impression memory in younger and older adults. *The Journal of Gerontology: Series B, 70*(5), 737–742. https://doi.org/10.1093/geronb/gbt132

Litt, M. D. (1988). Self-efficacy and perceived control: Cognitive mediators of pain tolerance. *Journal of Personality and Social Psychology, 54*, 149–160.

Matzen, L. E., Trumbo, M. C., Leach, R. C., & Leshikar, E. D. (2015). Effects of non-invasive brain stimulation on associative memory. *Brain Research, 1624*(3), 286–296. https://doi.org/10.1016/j.brainres.2015.07.036

McCurdy, M. P., Leach, R. C., & Leshikar, E. D. (2017). The generation effect revisited: Fewer constraints enhances item and context memory. *Journal of Memory and Language, 92*, 202–216. https://doi.org/10.1016/j.jml.2016.06.007

McCurdy, M. P., Leach, R., & Leshikar, E. D. (2019). Fewer constraints influence generation effect for source memory in younger, but not older adults. *Open Psychology, 1*, 168–184. https://doi.org/10.1515/psych-2018-0012

McCurdy, M. P., Frankenstein, A. N, Sklenar, A. M., & Leshikar, E. D. (2020a). Fewer generation constraints increase the generation effect for item and source memory through enhanced relational processing. *Memory, 28*(5). https://doi.org/10.1080/09658211.2020.1749283

McCurdy, M. P., Viechtbauer, W., Sklenar, A. M., Frankenstein, A. N., & Leshikar, E. D. (2020b). Theories of the generation effect and the impact of generation constraint: A meta-analytic review. *Psychonomic Bulletin & Review, 27*(6), 1139–1165. https://doi.org/10.3758/s13423-020-01762-3

McCurdy, M. P., Frankenstein, A. N., Sklenar, A. M., Urban Levy, P., & Leshikar, E. D. (2021). Examining the relationship between generation constraint and memory. *Memory & Cognition, 49*(4), 675–691. https://doi.org/10.3758/s13421-020-01119-0

McCurdy, M. P., & Leshikar, E. D. (in press). Contextual framework of the generation effect. *American Journal of Psychology.*

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19*(4/5), 494–513.

McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 371–385.

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in middle school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology, 27*, 360–372. https://doi.org/10.1002/acp.2914

Meyer, Z., McCurdy, M. P., Leach, R. C., Thomas, A. K., & Leshikar, E. D. (2020). Effects of survival processing on item and context memory: Enhanced memory for survival-relevant details. Frontiers in Psychology. *Frontiers in Psychology, 11*, 1–11. https://doi.org/10.3389/fpsyg.2020.02244

Miller, T. M., & Geraci, L. (2014). Improving metacognitive accuracy: How failing to retrieve practice items reduces overconfidence. *Consciousness and Cognition: An International Journal, 29*, 131–140.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519–533.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30–38.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*, 543–578.

Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology, 20*, 426–443.

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin, 144*(7), 710–756. https://doi.org/10.1037/bul0000151

Patel, S. P., McCurdy, M. P., Frankenstein, A. N., Sklenar, A. M., Urban Levy, P., Szpunar, K. K., & Leshikar, E. D. (in press). The reciprocal relationship between episodic memory and future thinking: How the outcome of predictions are subsequently memory. *Brain and Behavior.*

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879–891.

Prussia, G. E., & Kinicki, A. J. (1996). A motivational investigation of group effectiveness using social-cognitive theory. *Journal of Applied Psychology, 81*(2), 187.

Revelle, W. (2020) *psych: Procedures for personality and psychological research* (Version 2.0.9) [Computer software]. Northwestern University. https://CRAN.R-project.org/package=psych

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137*(4), 615–625.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130*(2), 261–288.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35*(9), 677.

Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher, 39*, 406–412.

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 233–239. https://doi.org/10.1037/a0017678

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.

RStudio Team. (2016). *RStudio: Integrated development for R*. RStudio.

Schunk, D. H. (1985). Self-efficacy and classroom learning. *Psychology in the Schools, 22*(2), 208–223. https://doi.org/10.1002/1520-6807(198504)22:2<208::AID-PITS2310220215>3.0.CO;2-7

Schunk, D. H., & Pajares, F. (2002). The development of academic self-efficacy. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation (pp. 15–31)*. Academic Press. https://doi.org/10.1016/B978-012750053-9/50003-6

Screbo, M. W., Warm, J. S., Dember, W. N., & Grasha, A. F. (1992). The role of time and cuing in a college lecture. *Contemporary Educational Psychology, 17*, 312–328.

Sklenar, A. M., Perez, J., McCurdy, M. P., Frankenstein, A. N., Motyl, M., & Leshikar, E. D. (2021). Person memory mechanism underlying approach and avoidance judgments of social targets. *Social Cognition, 39*, 747–772. https://doi.org/10.1521/soco.2021.39.6.747

Spreen, O., & Benton, A. L. (1977). *Neurosensory center comprehensive examination for aphasia*. University of Victoria, Neuropsychology Laboratory.

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin, 121*(3), 371.

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

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

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software, 59*(5), 1–38 http://www.jstatsoft.org/v59/i05/

Udeogu, O. J., Frankenstein, A. N., Sklenar, A. M., Urban Levy, P., & Leshikar, E. D. (in press). Predicting and remembering the behaviors of social targets: How prediction accuracy affects subsequent memory. *BMC Psychology.* https://doi.org/10.1186/s40359-022-00801-z

Vesonder, G. T., & Voss, J. F. (1985). On the ability to predict one's own responses while learning. *Journal of Memory and Language, 24*(3), 363–376.

Villasenor, J., Sklenar, A. M., Frankenstein, A. N., Urban Levy, P., McCurdy, M. P., & Leshikar, E. D. (2021). Value directed memory effects on item and context memory. *Memory & Cognition.* https://doi.org/10.3758/s13421-021-01153-6

Wechsler, D. (1981). *WAIS-R: Wechsler adult intelligence scale-revised*. Psychological Corporation.

Wong, S., Irish, M., Leshikar, E. D., Duarte, A., Bertoux, M., Savage, M., Hodges, J. R., Piguet, O., & Hornberger, M. (2017). The self-reference effect in dementia: Differential involvement of cortical midline structures in Alzheimer's disease and behavioural-variant frontotemporal dementia. *Cortex, 91*, 169–185. https://doi.org/10.1016/j.cortex.2016.09.013


**Open practices statement**

Materials and analysis code are available by emailing the corresponding author. This investigation was not preregistered.