

## Low-Stakes Quizzes Improve Learning and Reduce Overconfidence in College Students

**Kevin L. Kenney**

Kansas State University

**Heather Bailey**

Kansas State University

*Abstract: Retrieval practice is a straightforward and effective way to improve student learning, and its efficacy has been demonstrated repeatedly in the laboratory and in the classroom. In the current study, we implemented retrieval practice in the form of daily reviews in the classroom. Students (N = 47) in a cognitive psychology course completed a daily review at the beginning of each class. These consisted of 2-4 questions that encouraged students to practice retrieving material covered in lectures from the previous week. Then at the end of the semester, students took a comprehensive final exam consisting of content that was either on a daily review, a unit exam, both or neither. We replicated previous work showing that retrieval practice improved memory. Specifically, we found that students performed significantly better on questions whose information had been covered on both a daily review and unit exam. However, student performance did not differ amongst items covered only on a daily review, a unit exam, or on neither. Additionally, we extended previous work and found that students were significantly less overconfident for information covered on both a daily review and unit exam. The current results indicate that retrieval practice helps college students remember material over the course of a semester and also improves their ability to evaluate their own knowledge of the material.*

*Keywords: retrieval practice, learning; memory, metacognition*

A major goal of educators is for their students to learn course information effectively and to retain the information over a long period of time. However, students often struggle to develop effective study strategies (Karpicke, Butler & Roediger, 2009; Wissman, Rawson & Pyc, 2012). One simple and effective method of improving long-term learning is retrieval practice: the process of attempting to retrieve information at some point after it was initially learned. For instance, teachers can initiate retrieval practice in the classroom by asking students questions about the material, giving pop quizzes, or reviewing for an exam. Students can also initiate the process while studying by attempting to answer review questions in a textbook, creating flashcards, or having classmates test them. Practice tests in each of these scenarios occur during the learning process, prior to a high-stakes exam. In other words, teachers and students can use tests as a study tool to improve learning, and not just as a form of assessment (for reviews, see Rawson & Dunlosky, 2011; Roediger & Butler, 2011).

Previous work has demonstrated that practice tests improve memory performance in both laboratory and classroom settings. Such memory improvements were observed for text materials (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Roediger & Karpicke, 2006), definitions (Pan & Rickard, 2017), foreign language words (Fritz, Morris, Acton, Voelkel & Etkind, 2007; Pyc & Rawson, 2007), general knowledge facts (e.g., Agarwal, Finley, Rose, & Roediger, 2016; Carpenter, Pashler, Wixted, & Vul, 2008), and spatial information (e.g., Carpenter & Kelly, 2012; Carpenter & Pashler, 2007). And while the simple practice of retrieving information benefits memory, providing students feedback (i.e., the correct answer) immediately after the practice test further increases its beneficial effects (Butler, Karpicke, & Roediger, 2008).

The benefits of retrieval practice are not merely due to being exposed to the material again. A wealth of research has shown that practice tests improve memory to a greater extent than just restudying the material (e.g., Carrier & Pashler, 1992; Roediger & Karpicke, 2006). Further, research has shown that the more often students correctly retrieve information (e.g., 3 times vs. 1 time) the more likely they are to remember the information on a final exam (e.g., Karpicke & Roediger, 2008; Pyc & Rawson, 2009). In sum, research has demonstrated that repeated retrieval practice is a powerful strategy to improve long-term memory.

Despite its benefits, students often do not use retrieval practice when studying on their own (Karpicke, Butler & Roediger, 2009; Wissman et al., 2012). Rather, they use techniques such as highlighting the text and rereading notes, which boost the familiarity of the material—giving students the illusion of learning—but do not necessarily improve their comprehension of the material. The consequences of such study strategies are that students consistently overestimate how well they understand the course material and how likely they will be able to remember it on an exam. In fact, students tend to be very overconfident, predicting that they would earn a grade up to 30% higher than they actually earned (Hacker, Bol, Horgan & Rakow, 2000). Unfortunately, this effect is even larger for the lowest performing students (e.g., Hacker et al., 2000; Keleman, Winningham & Weaver, 2007; Krueger & Mueller, 2002; Nietfeld, Cao & Osborne, 2005).

When students are overconfident, they overestimate how much they have learned and prematurely decide to stop studying the material, which results in less learning (Dunlosky & Rawson, 2012). Such choices can drastically hurt their performance in the classroom. One explanation for students' overconfidence is the failure to accurately monitor their learning, which is an important *metacognitive process*. Highlighting text, for instance, does not encourage students to metacognitively monitor their learning to determine whether or not the information is well understood.

Imagine that a student is rereading a chapter in the textbook in preparation for a psychology exam and they come across the term *confirmation bias*. It seems familiar because they remember hearing about it in class and seeing it in their notes. Due to this feeling of familiarity, the student believes that they have learned the term well enough and moves on to another section. Unfortunately, the student did not evaluate their memory and understanding for the material, and as a result, they may not be able to correctly retrieve information related to confirmation bias on the exam. If the student had instead tested their memory for the term (e.g., by using flashcards), then they would have been able to assess how well they knew the information. If they failed to retrieve the information on the flashcard, then they could put that card at the back of the deck and try again sometime later.

Thus, retrieval practice should reduce students' overconfidence for class material. In fact, some work conducted in a laboratory setting demonstrated that retrieval practice improves some aspects of students' ability to assess their own learning. For instance, Pyc and Rawson (2012) found that the more often students correctly retrieve information, the more accurately they predict their exam performance. Such results could have important implications for student learning because the more they test their memory for course material, the better they should be able to evaluate which information they have and have not learned well. Students could then use these evaluations to make more effective study decisions by spending more time reviewing the information that was not well learned.

Another important variable to consider is how these retrieval attempts are spaced in time. That is, long-term memory is much better when students spread out their study sessions as compared to when they mass them (i.e., "cram"; Hintzman, 1974; Peterson, Wampler, Kirkpatrick & Saltzman, 1963; for a review, see Maddox, 2016). Combining repeated retrieval attempts with appropriate spacing is one of the most potent strategies for improving long-term memory (Cull, 2000; Landauer & Eldridge, 1967; Pyc & Rawson, 2009).

Importantly, research has demonstrated the efficacy of retrieval practice in the classroom. Retrieval practice with feedback has boosted memory for preschool aged children (Lipowski, Pyc, Dunlosky, & Rawson, 2014), middle school students (e.g., McDaniel, Agarwal, Huelser, McDermott & Roediger, 2011; McDaniel, Thomas, Agarwal, McDermott & Roediger, 2013; McDermott, Agarwal, D'Antonio, Roediger & McDaniel, 2014; Roediger, Agarwal, McDaniel & McDermott, 2011), high school students (McDermott et al, 2014), college students (Cranney, Ahn, McKinnon, Morris & Watts, 2009; Lyle & Crawford, 2011; McDaniel, Anderson, Derbish & Morrisette, 2007), and medical students (Larsen, Butler & Roediger, 2009). Research has shown that students who engage in retrieval practice throughout the academic year have higher exam grades (McDaniel et al, 2011; Roediger et al, 2011) and higher final course grades (Leeming, 2002) as compared to students who do not engage in repeated retrieval practice.

### Current Study

We implemented retrieval practice in the classroom through a technique we call *daily reviews*. At the beginning of each class, students responded to between two and four review questions from material covered in lectures and assigned readings from the previous week. They were encouraged to attempt to answer the questions without looking at their notes to help them assess how well they had learned the information. However, they were allowed to use their notes and textbook, if necessary. This technique differs from traditional “pop quizzes” because student responses were *not* graded for content – rather they received a completion grade. The goal of this assessment was to encourage students to practice retrieving information and evaluating how well they had learned it in a low-stakes situation, while at the same time reinforcing course content (e.g., encoding, retrieval, episodic long-term memory, metacognition). After collecting responses, we provided the correct answers so that students had the opportunity to relearn the information if it were recalled incorrectly. In addition, we always allowed for questions as to why certain responses were correct or incorrect. We used a mixture of definitional and application questions on the daily reviews so that students could practice retrieving the information in different ways and because prior work has shown that practicing application questions can improve exam performance on both question types (McDaniel et al., 2013).

The main goal of the current study was to assess the efficacy of retrieval practice in our Cognitive Psychology course. We evaluated whether retrieval practice improves long-term retention and, most importantly, metacognitive skills (as measured by confidence judgments) in our students. To do so, we manipulated questions on the comprehensive final exam such that they either had appeared on a daily review, a unit exam, both or neither. We also collected confidence ratings for each response on the final exam.

Based on the robust success of retrieval practice in the literature, we predicted that information would be better remembered on the final exam if it had appeared on a unit exam or daily review since these events are forms of retrieval practice. Further, given that retrieval practice seems to be dose-dependent (e.g., Karpicke & Roediger, 2008; Pyc & Rawson, 2009), we also predicted that memory would be best for information that appeared on both a daily review and a unit exam. Most importantly, we predicted that having students engage in retrieval practice (on a daily review, unit exam, or both) and assess how well the material was learned would result in better metacognitive skills, as measured by more accurate confidence ratings, for previously-tested material.

## Method

### *Design*

In this one-way within-subjects quasi-experimental design, question condition was a within-subjects variable with four levels: *Daily Review Only*, *Exam Only*, *Daily Review + Exam*, and *None*. We also had three dependent variables: Final Exam Performance, Confidence Judgment Magnitude, and Confidence Judgment Accuracy.

### *Participants*

The class consisted of 47 students enrolled in an upper-level cognitive psychology course, 45 (40 women, 5 men) of whom completed the final exam and 44 (39 women, 5 men) of whom opted in to providing confidence ratings. All 45 who completed the exam were included in performance analyses, and all 44 who provided confidence ratings were included in the confidence analyses. Enrolled students consisted of 1 first-year, 2 sophomores, 25 juniors, and 16 seniors (age of participants was not collected as a part of this project). Recent meta-analyses evaluating the effect of retrieval practice on memory reported medium to strong effect sizes (Adesope, Trevisan, & Sundararajan, 2017:  $g = .74$ ; Rowland, 2014:  $g = .55$ ). Using the most conservative estimate of effect size ( $g = 0.55$ ), we conducted a power analysis in G\*Power 3.1.7. Based on a one-tailed hypothesis, with an effect size of 0.55, alpha = .05 and power of .95, G\*Power indicated that a total sample size of 38 would be needed to detect the within-subjects effect of retrieval practice on memory. Thus, our sample size should be sufficient to detect the effects of interest. This study was approved as exempt research; thus, students did not need to provide informed consent. Students received no additional compensation beyond credit awarded for completing course exams, final exam, and daily reviews exercised (described in detail in the materials/procedure section).

### *Materials and Procedure*

Students enrolled in cognitive psychology were instructed in a typical university lecture format that was delivered twice per week for 75 minutes each day. Students completed four regular unit exams (of which the lowest one score was “dropped” and did not count towards the course grade; optionally, one exam could be skipped if the student chose to do so), a cumulative final exam, and 20 daily review assignments. All exams included approximately 30 multiple-choice and 15 short answer (matching, fill-in-the blank, single-sentence, or short-essay) questions. Some of the questions on each exam were related to questions on daily review assignments and the rest had been discussed in lecture and/or readings but had not been explicitly reviewed in class. After grading the unit exams, students had the opportunity to review their responses as well as our feedback in class; however, they were not allowed to take the exams home. The same is true for the daily reviews. All students had the opportunity to come to the professor’s office at any point in the semester to review their exams and daily review responses.

*Daily reviews.* The daily review assignments consisted of two to four short answer questions presented at the beginning of lecture that encouraged students to practice retrieving material covered in lectures from the previous week (e.g., “In what ways have animals shown evidence of language comprehension”, “Which type of reasoning involves making conclusions that are probably true”, “If you were to see a picture of Kansas State University’s football field, which part of your brain is highly specialized to respond?”). Students were encouraged to answer the questions without using their notes

or textbook. Although the use of their notes and textbook was not prohibited, the students were aware of the benefits of retrieval practice.

We covered the relevant research early in the semester in the *Encoding and Retrieval* module: specifically, that testing oneself (e.g., with flashcards and daily reviews) promotes learning above and beyond simply re-reading one's textbook and notes. In this module, we also discussed how self-testing could help students develop their metacognitive skills by helping them identify which information has been well learned and which information has not. We also discussed how the most effective study strategy is to focus future study time on the information that is not well learned (i.e., not correctly recalled on a daily review and/or on flashcards).

The students were allowed approximately three to five minutes to complete the daily review. These assignments were awarded points for completion, not correctness. Ten of these 20 assignments were randomly-selected (using the =rand() function in Microsoft® Excel) to be awarded one extra credit point if students answered all questions correctly (students were not made aware of which daily review assignments would be evaluated for extra credit, but knew that 10 throughout the semester would be selected). This extra credit incentive was offered to further encourage students to practice retrieving the information in an effortful manner.

*Cumulative final exam.* The final exam consisted of 52 total items (34 multiple-choice, 18 short answer). To assess the efficacy of retrieval practice on long-term retention in a classroom setting, final exam questions were coded as having been on a daily review assignment only (Daily Review Only; 8 items), having been on an exam only (Exam Only; 15 items), having been on both an exam and a daily review assignment (Daily Review + Exam; 13 items), or having been on neither an exam nor a daily review assignment (None; 16 items). Importantly, we never tested the exact question twice. That is, final exam questions that contained content from a daily review (Daily Review Only and Daily Review + Exam items) shared the same content information but were either changed at the surface level ( $n = 5$ ) or at the conceptual level ( $n = 18$ ). An example of a surface level change is “When solving a problem, novices tend to use surface characteristics and experts tend to use \_\_\_\_” (daily review) and “Experts categorize problems based on \_\_\_\_” (final exam example). An example of a conceptual change is “Do we attend to irrelevant stimuli during a low-load or high load task?” (daily review example) and “Chelsea is sitting in her Cognitive Psychology class. She is least likely to see her friend waving from the hallway when she is completing which task?” (final exam example).

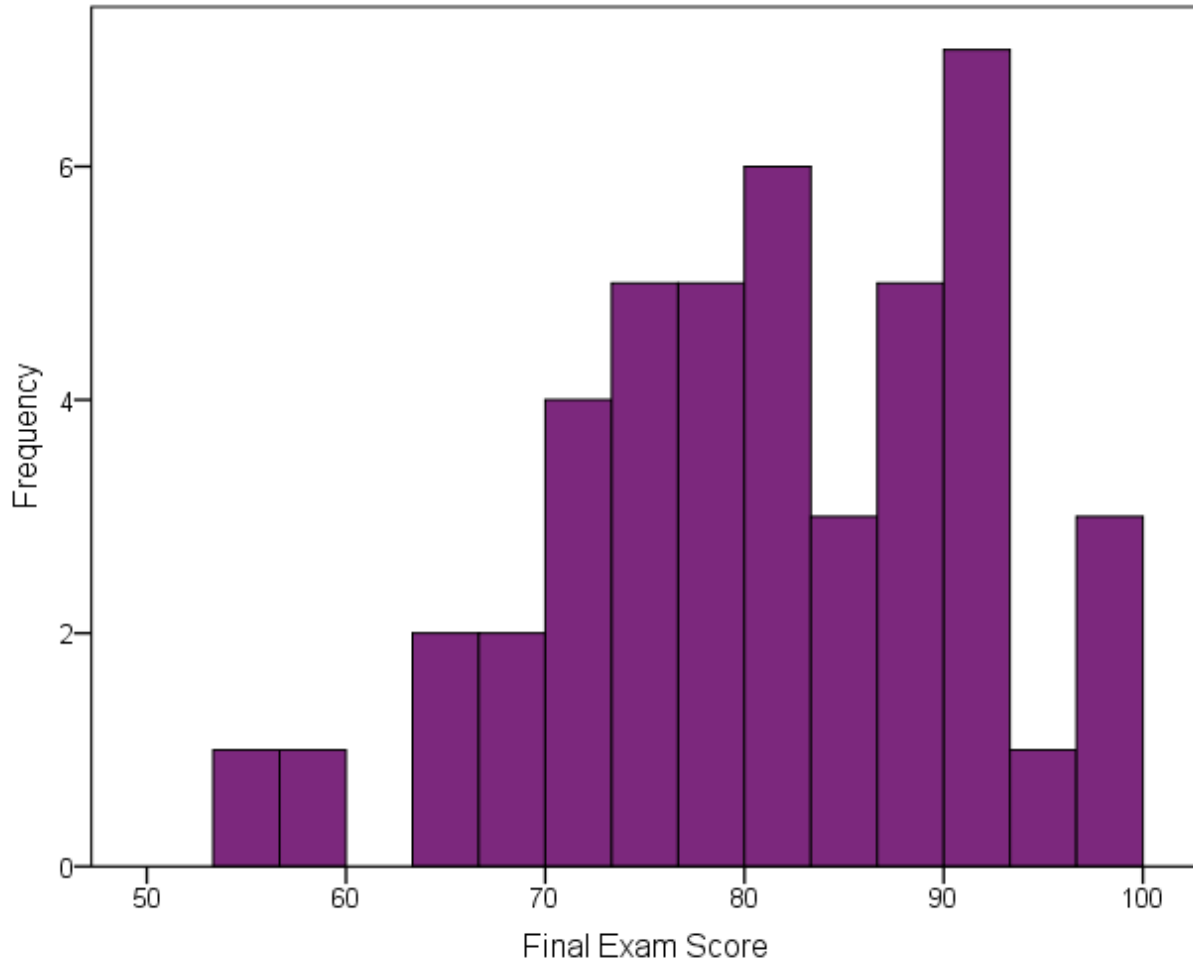
Performance differences between question conditions (Daily Review Only, Daily Review + Exam, Exam Only, and None) were tested using repeated-measures ANOVAs. After finishing the final exam, students completed a set of confidence ratings. For each question on the final exam, students rated (on a 0-100 scale) how confident they were that their answer was correct. Note that the confidence ratings were collected after students completed the exam, but they still had access to the questions and their responses while making their ratings.

## Results

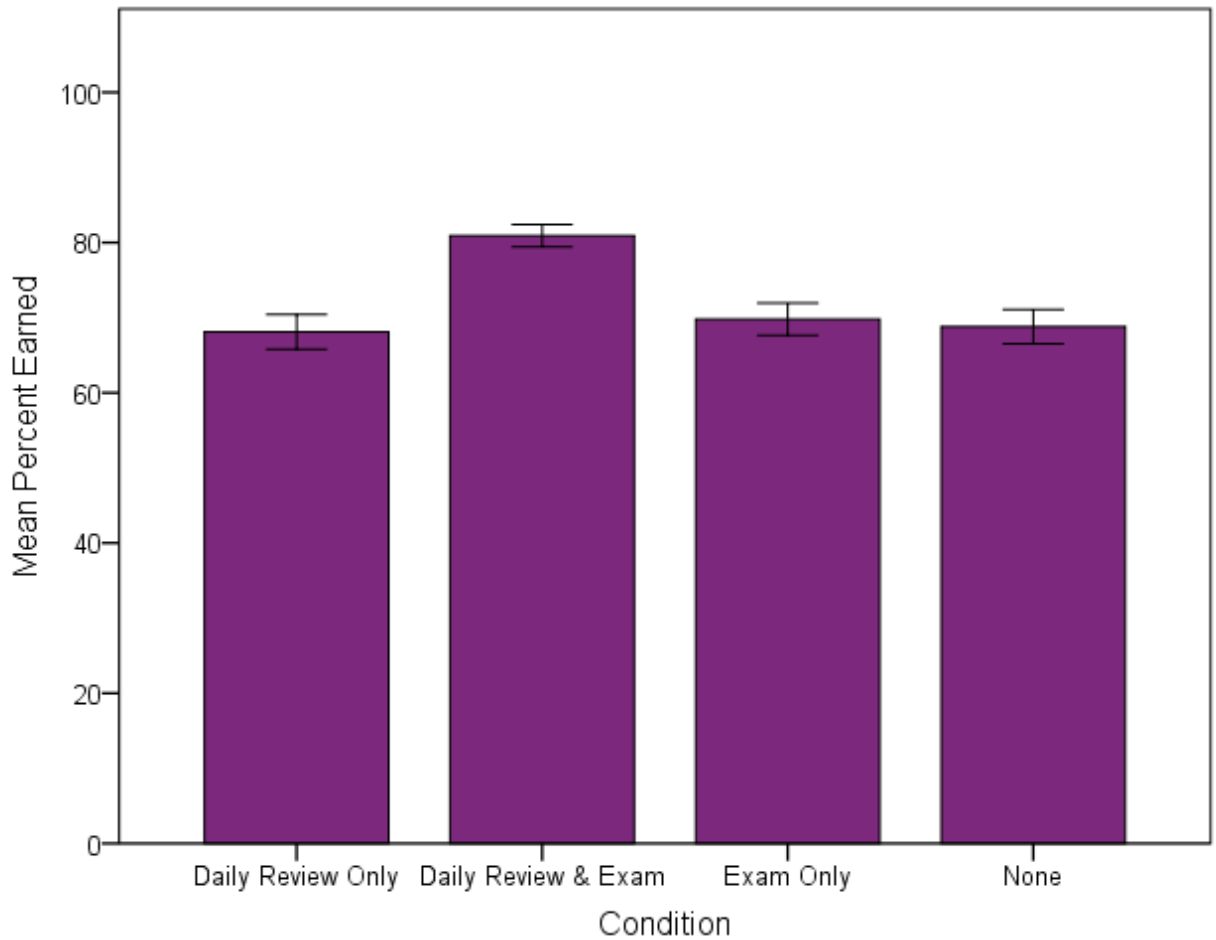
### *Final Exam Performance*

The average percentage earned on the final exam was 81.08% ( $SD = 10.41$ ). A histogram of final exam scores is provided in Figure 1 and mean final exam performance is reported by condition in Figure 2. To test the efficacy of daily review assignments on cumulative final exam performance, we performed a repeated measures ANOVA. We found a significant effect of question condition on final exam performance,  $F(3, 132) = 13.89$ ,  $p < .001$ , partial  $\eta^2 = .24$ . We then ran post hoc analyses, with Bonferroni correction, and found that performance was significantly higher for Daily Review + Exam

( $M = 80.90$ ,  $SD = 9.99$ ) compared to Daily Review Only ( $M = 68.11$ ,  $SD = 15.52$ ), E ( $M = 69.81$ ,  $SD = 14.45$ ), and None ( $M = 68.81$ ,  $SD = 15.29$ ). No significant differences between any other conditions were detected. Thus, we replicated prior work showing that retrieval practice, especially multiple retrieval attempts, improve long-term memory (e.g., Karpicke & Roediger, 2008; Pyc & Rawson, 2009).



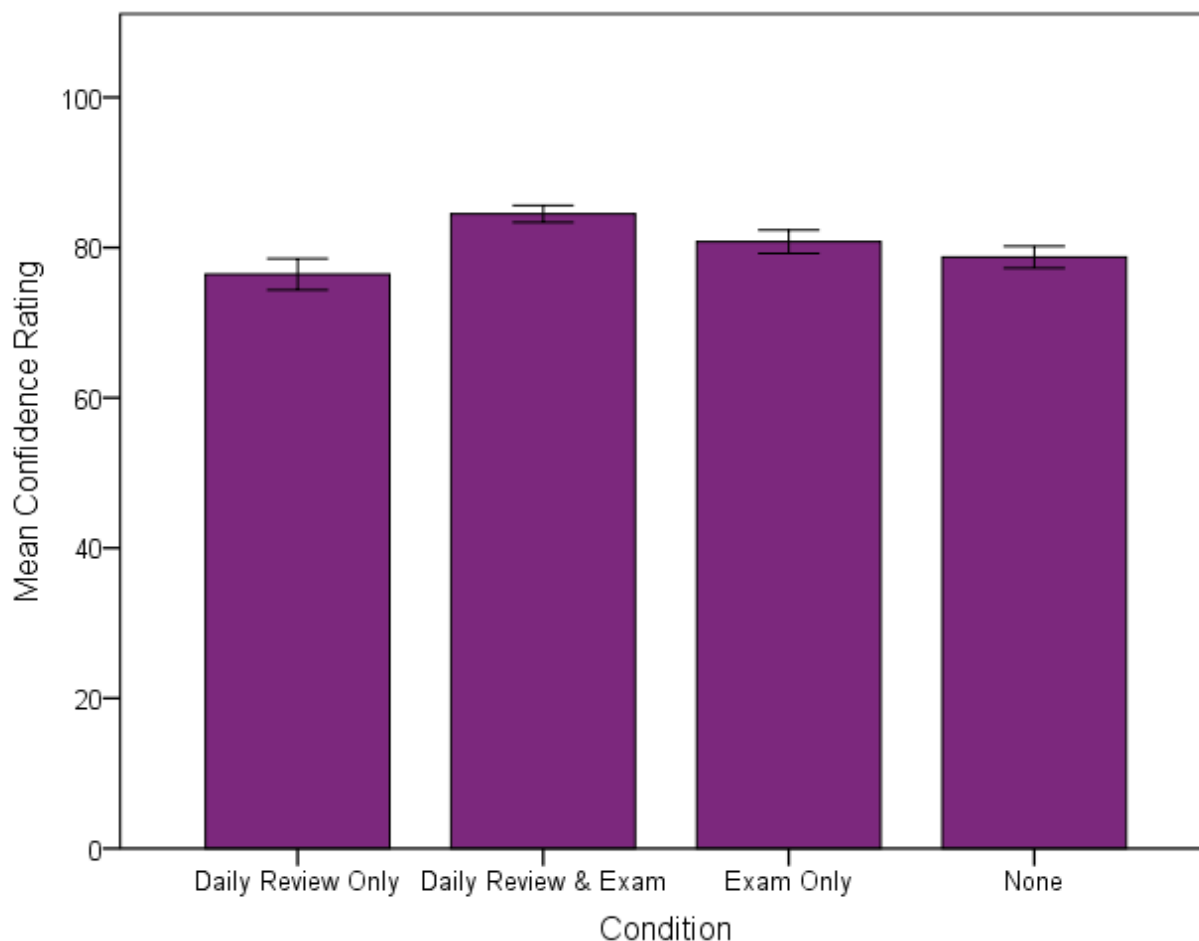
**Figure 1. Distribution of final exam scores.** The distribution had a non-significant negative skew ( $-0.45$ ,  $SE = 0.35$ ) and non-significant kurtosis ( $-0.14$ ,  $SE = 0.70$ ).



**Figure 2. Mean percentage earned by question condition.** Error bars represent one standard error.

### *Confidence Ratings*

Now that we replicated the effect of retrieval practice on memory, we turn to the novel effect of interest: the effect of retrieval practice on reducing student overconfidence. To do so, we evaluated confidence ratings, which were on a scale of 0-100 (0 represented no confidence that their response was correct and 100 indicated absolute confidence that it was correct). The mean confidence rating across the entire exam was 81.0 ( $SD = 23.0$ ) and is presented by question condition in Figure 3. We ran two sets of analyses on confidence ratings: one evaluating the magnitude of confidence ratings and one evaluating the accuracy.

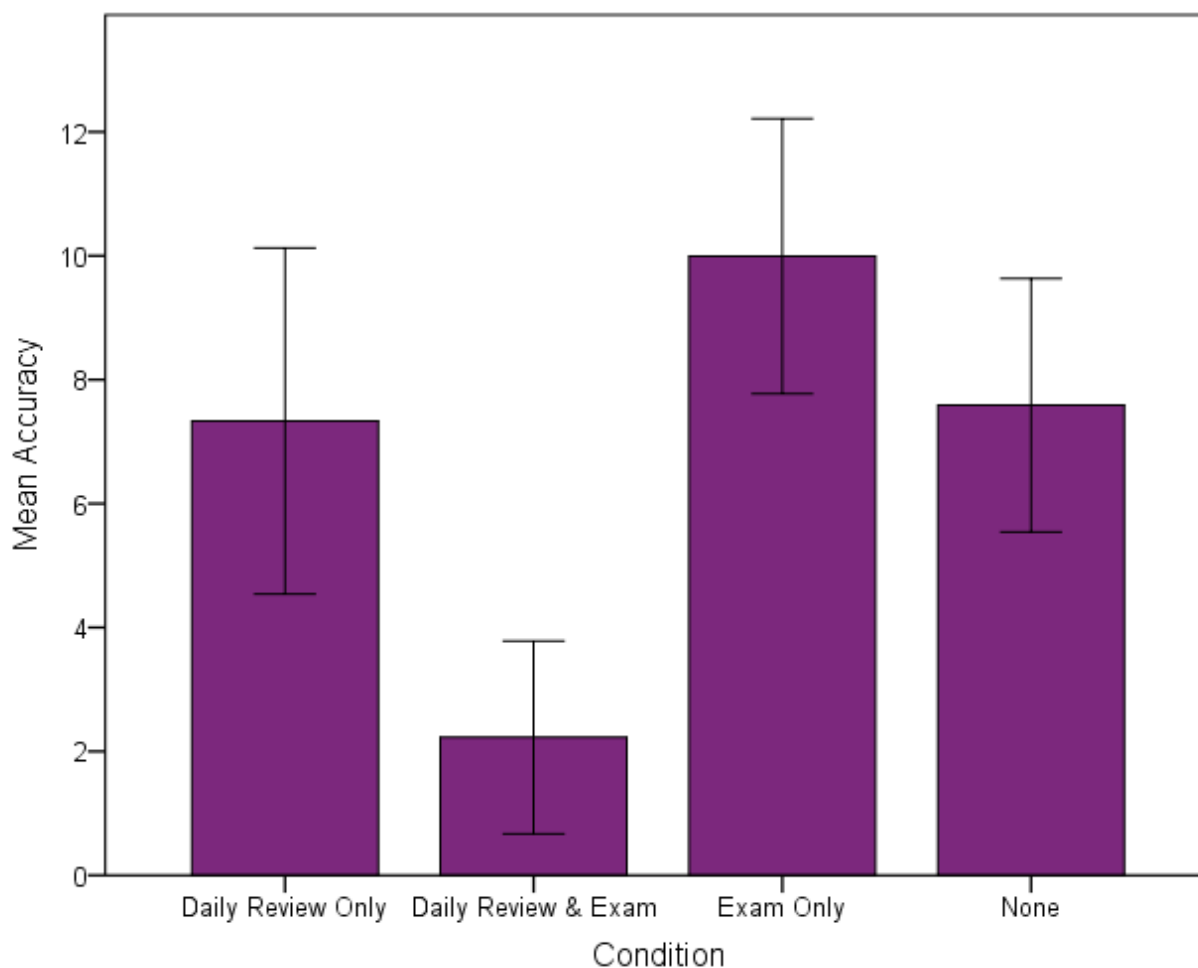


**Figure 3. Confidence ratings by question condition.** Error bars represent one standard error.

*Confidence Ratings Magnitude.* Again, we ran a repeated-measures ANOVA and found that confidence ratings differed significantly based on question condition,  $F(3, 129) = 12.89, p < .001$ , partial  $\eta^2 = .23$ . Then we ran post hoc analyses, with Bonferroni correction, and found that confidence ratings were significantly higher for Daily Review + Exam ( $M = 84.48, SD = 7.50$ ) compared to Daily Review Only ( $M = 76.44, SD = 13.78$ ), Exam Only ( $M = 80.78, SD = 10.30$ ), and None ( $M = 78.74, SD = 9.69$ ). No significant differences between any other conditions were detected.

*Confidence Ratings Accuracy.* Next, we calculated the accuracy of confidence ratings by subtracting the percentage of points earned on that item from the confidence rating on an item. Thus, positive values can be interpreted as overconfidence, negative values as under confidence, and values closer to zero as more accurate. As shown in Figure 4, students were consistently overconfident, regardless of question condition; however, we ran a repeated-measures ANOVA and found that students' confidence rating accuracy significantly differed by question condition,  $F(3, 129) = 3.68, p = .014$ , partial  $\eta^2 = .08$ . We then ran post hoc analyses, with Bonferroni correction, and found that confidence rating accuracy was significantly better (i.e., the difference between confidence rating and percentage earned was closer to 0) for Daily Review + Exam ( $M = 2.23, SD = 10.33$ ) compared to Exam Only ( $M = 9.99, SD = 14.73$ ), but we detected no other differences between question conditions [Daily Review Only ( $M = 7.33, SD = 18.54$ ); None ( $M = 7.59, SD = 13.57$ )].





**Figure 4. Mean confidence rating accuracy by question condition.** Error bars represent one standard error.

## Discussion

Our main goal for the current study was to evaluate whether retrieval practice (on a daily review or a unit exam) improves students' performance and metacognitive skills on a cumulative final exam. We found that retrieval practice did improve final exam performance and metacognitive accuracy, but only when the information was presented on *both* a daily review assignment and a unit exam. Surprisingly, in terms of performance and confidence judgments, items presented on only a daily review assignment and those presented only a unit exam were no different than the control items.

One possible explanation is that more than one test is necessary to improve learning over such long retention intervals. In the current study, time between initial learning and the final exam ranged from approximately 2 weeks to 3.5 months. Prior work has demonstrated that optimal learning occurs when *multiple* tests are spaced in time, especially when there is a large gap between learning and final test (e.g., Cepeda, Coburn, Rohrer, Wixted, Mozer & Pashler, 2006). Thus, our most optimal condition in the current study was when information was retrieved at least twice (on a daily review & unit exam) and these retrievals were distributed in time. That is, students' final exam performance did benefit when the material was reviewed in class more than one time during the semester.

Further, students were less overconfident when information was presented on both a daily review assignment and exam. Retrieval practice provided students with more information about their learning—whether or not the information was well learned—which, in turn, improved their metacognitive awareness. When students more effectively monitor their learning, their confidence ratings should be more accurate, as we observed in the current study. Previous research has demonstrated that a metacognitive monitoring intervention, in which students frequently assessed their own learning after class, improved confidence accuracy and exam performance (Nietfeld, Cao & Osborne, 2006). Here, we observed that retrieval practice produced similar results.

Such effects dovetail nicely with the prior work evaluating repeated retrieval and spaced retrieval. Our current results replicate prior work demonstrating that retrieval practice benefits memory (e.g., Agarwal et al., 2008; 2017; Butler et al, 2008; Carpenter et al., 2008; Carpenter & Kelly, 2012; Carpenter & Pashler, 2007; Cranney et al., 2009; Lyle & Crawford, 2011; McDaniel et al., 2007; Pan & Rickard, 2017; Pyc & Rawson, 2007; Roediger & Karpicke, 2006). Further, these benefits are even larger when the repeated retrievals are spaced apart (Cull, 2000; Landauer & Eldridge, 1967; Pyc & Rawson, 2009). In our class, the retrieval attempts on the daily reviews, the unit exams and the final exam were spaced out by weeks to months.

Laboratory-based research has also demonstrated that retrieval practice exerts its effects not only on memory, but also on students' metacognitive skills. That is, rather than relying on their feelings of familiarity (i.e., the illusion of learning), retrieval practice forces students to evaluate their memory for the material. If they are able to correctly recall the information, their confidence in their learning of that material should increase. If they are unable to correctly recall the information, their confidence should be reduced and, in turn, the students can make a strategic choice to spend more time on that material.

It is possible that the effects we observed here are not necessarily due to repeated retrievals but instead due to increased salience of the material. In other words, information that was covered on a daily review and a unit exam may be perceived as more important to the professor as compared to other material and, therefore, students may have spent more time studying this material outside of class. By this logic, though, we would have expected students to recall information tested once (Daily Review Only or E items) better than information never tested (None items). We believe that our effects are driven by repeated retrievals given the wealth of support for its memorial benefits; however, future work should attempt to disentangle the effects of retrieval practice and salience in the classroom.

One caveat that we should note here is that these results only accounted for instructor-initiated retrieval practice. We assessed memory performance and confidence ratings based on information that had been reviewed or tested in the classroom, but we had no way of knowing the extent to which the students initiated retrieval practice on their own. Anecdotally, several students reported using flashcards or quizzing a peer, but we do not know which material or how many times they practiced retrieving it. However, we believe this actually bolsters the current results. Despite any noise in the final exam performance due to individual differences in students' study strategies at home, we found that instructor-initiated retrieval practice still had a significant impact on exam performance and students' confidence.

However, the current data suggest that students need to retrieve the information *at least* two times to see the benefits described above. In the current study, this entailed retrieving the information on a daily review and on a unit exam. Future research in a classroom setting is needed to determine whether the same benefits of these two practice retrievals could arise from two daily reviews alone or if the process retrieving on a formal unit exam carries more of the benefit. Further, future research could evaluate the benefits of more than two retrieval attempts. Work from a laboratory setting has

already shown that students can benefit even further from multiple retrieval attempts but with diminishing returns (Pyc & Rawson, 2012).

In addition, we discussed with the students *why* they take the daily reviews. They learned about the benefits of retrieval practice; specifically, that testing oneself improves learning and memory. We also discussed how retrieval practice is more effective than simply re-reading the textbook and notes. The daily review assignments align well with a learning goal that is important for most instructors and students—to learn basic concepts and demonstrate that knowledge. Explaining to students the research on retrieval practice and its connection to learning goals can help them take ownership over their own learning.

One of the major advantages of this low-stakes quizzing approach is that it is easy to implement. It helps instructors to better assess what students know and do not know well. They can then use performance on these quizzes to determine which information should be covered in more detail or presented in a different manner to optimize student learning. Further, these quizzes can be implemented across a broad range of class types (e.g., lecture, seminar), across a broad range of disciplines (e.g., psychology, kinesiology, physics), and across many different situations. For instance, this approach can help students prepare for exams, retain information from one class to another or apply their knowledge in a real-world setting.

### *Conclusion*

Overall, we demonstrated that students' cumulative final exam performance and metacognitive skills were better for information that had been tested throughout the semester. Such results add to the growing literature on the benefits of retrieval practice. Specifically, we demonstrated that a simple technique, such as reviewing material from previous lectures in a low-stakes manner, can help students retain material over the course of a semester and can also improve their ability to metacognitively evaluate their own knowledge.

### **Acknowledgements**

Thanks to Kristen McGatlin for help with data scoring and data entry.

### **References**

- Agarwal, P. K., Finley, J. R., Rose, N. S., & Roediger III, H. L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory*, 25, 764-771. <https://doi.org/10.1080/09658211.2016.1220579>
- Agarwal, P. K., Karpicke, J. D., Kange, S. H. K., Roediger III, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, 22, 861-876. <https://doi.org/10.1002/acp.1391>
- 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, 659-701. <https://doi.org/10.3102/0034654316689306>
- Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 918-928. <https://doi.org/10.1037/0278-7393.34.4.918>

- Carpenter, S. K., & Kelly, J. W. (2012). Tests enhance retention and transfer of spatial learning. *Psychonomic Bulletin & Review*, *19*, 443-448. <https://doi.org/10.3758/s13423-012-0221-2>
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36*, 438-448. <https://doi.org/10.3758/mc.36.2.438>
- Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, *14*, 474-478. <https://doi.org/10.3758/bf03194092>
- Carrier, M. & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*, 632–642. <https://doi.org/10.3758/bf03202713>
- Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, *56*, 236–246. <https://doi.org/10.1027/1618-3169.56.4.236>
- Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. *European Journal of Cognitive Psychology*, *21*, 919-940. <https://doi.org/10.1080/09541440802413505>
- Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, *14*, 215–235. [https://doi.org/10.1002/\(sici\)1099-0720\(200005/06\)14:3<215::aid-acp640>3.3.co;2-t](https://doi.org/10.1002/(sici)1099-0720(200005/06)14:3<215::aid-acp640>3.3.co;2-t)
- Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction*, *22*, 271-280. <https://doi.org/10.1016/j.learninstruc.2011.08.003>
- Fritz, C., Morris, P. E., Acton, M., Voelkel, A. R., & Etkind, R. (2007). Comparing and combining expanding retrieval practice and the keyword mnemonic for foreign vocabulary learning. *Applied Cognitive Psychology*, *21*, 499-526. <https://doi.org/10.1002/acp.1287>
- Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, *92*, 160-170. <https://doi.org/10.1037/0022-0663.92.1.160>
- Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in Cognitive Psychology: The Loyola Symposium* (pp. 77–99). Potomac, MD: Erlbaum.
- Karpicke, J. D., Butler, A. C. & Roediger III, H. L. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory*, *17*, 471-479. <https://doi.org/10.1080/09658210802647009>
- Karpicke, J. D., & Roediger III, H.L. (2008). The critical importance of retrieval for learning. *Science*, *15*, 966-968. <https://doi.org/10.1126/science.1152408>
- Kelemen, W. L., Winningham, R. G., & Weaver, C. A., III. (2007). Repeated testing sessions and scholastic aptitude in college students' metacognitive accuracy. *European Journal of Cognitive Psychology*, *19*, 689–717. <https://doi.org/10.1080/09541440701326170>
- Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology*, *82*, 180–188. <https://doi.org/10.1037/0022-3514.82.2.180>  
<https://doi.org/10.1111/j.1365-2923.2009.03518.x>
- Landauer, T. K., & Eldridge, L. (1967). Effects of tests without feedback and presentation-test interval in paired-associate learning. *Journal of Experimental Psychology*, *75*, 290–298. <https://doi.org/10.1037/h0025047>
- Larsen, D. P., Butler, A. C., & Roediger III, H. L. (2009). Repeated testing improves long-term retention relative to repeated study: A randomized controlled trial. *Medical Education*, *43*,

- Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology, 29*, 210-212. [https://doi.org/10.1207/s15328023top2903\\_06](https://doi.org/10.1207/s15328023top2903_06)
- Lipowski, S. L., Pyc, M. A., Dunlosky, J., & Rawson, K. A. (2014). Establishing and explaining the testing effect in free recall for young children. *Developmental Psychology, 50*, 994-1000. <https://doi.org/10.1037/a0035202>
- Lyle, K. B., & Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. *Teaching of Psychology, 38*, 94-97. <https://doi.org/10.1177/0098628311401587>
- Maddox, G. B. (2016). Understanding the underlying mechanism of the spacing effect in verbal learning: A case for encoding variability and study-phase retrieval. *Journal of Cognitive Psychology, 28*, 684-706. <https://doi.org/10.1080/20445911.2016.1181637>
- McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger III, H. L. (2011). Test enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology, 103*, 399-414. <https://doi.org/10.1037/a0021782>
- 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*, 494-513. <https://doi.org/10.1080/09541440701326154>
- McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger III, H. L. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology, 27*, 360-372. <https://doi.org/10.1080/09541440701326154>
- McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger III, H. L., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied, 20*, 3-21. <https://doi.org/10.1037/xap0000004>
- Miller, T. M., & Geraci, L. (2014). Improving metacognitive accuracy: How failing to retrieve practice items reduces overconfidence. *Consciousness and Cognition, 29*, 131-140. <https://doi.org/10.1016/j.concog.2014.08.008>
- Nietfeld, J. L., Cao, L., & Osborne, J. W. (2005). Metacognitive monitoring accuracy and student performance in a postsecondary classroom. *The Journal of Experimental Education, 74*, 7-28.
- Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning, 1*, 159-179. <https://doi.org/10.1007/s10409-006-9595-6>
- Pan, S. C., & Rickard, T. C. (2017). Does retrieval practice enhance learning and transfer relative to restudy for term-definition facts? *Journal of Experimental Psychology: Applied, 23*, 278-292. <https://doi.org/10.1037/xap0000124>
- Peterson, L.R., Wampler, R., Kirkpatrick, M., & Saltzman, D. (1963). Effect of spacing of presentations on retention of paired-associates over short intervals. *Journal of Experimental Psychology, 66*, 206-209. <https://doi.org/10.1037/h0046694>
- Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition, 35*, 1917-1927. <https://doi.org/10.3758/bf03192925>
- Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*, 437-447. <https://doi.org/10.1016/j.jml.2009.01.004>
- Pyc, M. A., & Rawson, K. A. (2012). Are judgments of learning made after correct responses during retrieval practice sensitive to lag and criterion level effects? *Memory & Cognition, 40*, 976-988. <https://doi.org/10.3758/s13421-012-0200-x>

- Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, *140*, 283–302. <https://doi.org/10.1037/a0023956>
- Roediger III, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *17*, 382-395. <https://doi.org/10.1037/a0026252>
- Roediger III, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*, 20-27. <https://doi.org/10.1016/j.tics.2010.09.003>
- Roediger III, H. L. & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological Bulletin*, *140*, 1432-1463. <https://doi.org/10.1037/a0037559>
- Wissman, K. T., Rawson, K. A. & Pyc, M. A. (2012). How and when do students use flashcards? *Memory*, *20*, 568-579. <https://doi.org/10.1080/09658211.2012.687052>