Title
Learning depends on knowledge: The benefits of retrieval practice vary for facts and skills

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Carvalho, Paulo
Rachatasumrit, Napol
Koedinger, Kenneth R

Publication Date
2022

Peer reviewed

Learning depends on knowledge:
The benefits of retrieval practice vary for facts and skills

Paulo F. Carvalho (pcarvalh@andrew.cmu.edu)
Human-Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Ave
Pittsburgh, PA 15213 USA

Napol Rachatasumrit (nrachata@andrew.cmu.edu)
Human-Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Ave
Pittsburgh, PA 15213 USA

Kenneth R. Koedinger (koedinger@cmu.edu)
Human-Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Ave
Pittsburgh, PA 15213 USA

Abstract

Retrieval practice of information through testing has been shown to improve learning. So has studying examples. In this paper, we address inconsistencies in the literature concerning which of these two approaches is best. We test the hypothesis that learning depends on what is being learned; whereas practice emphasizes memorization, studying examples allows for selectivity of encoding, resulting in different information being learned. Accordingly, we predicted that practice will improve learning in situations that emphasize memorization (such as learning facts or simple associations), whereas studying examples will improve learning in situations where there are multiple pieces of information available and selectivity is necessary (such as when learning skills or procedures). We report evidence from a laboratory study using naturalistic materials showing results consistent with these predictions.

Keywords: practice; retrieval; study; worked-examples

Introduction

When repeated studying of the same materials is replaced by testing or retrieval practice, memory and learning are enhanced (Roediger & Karpicke, 2006; Roediger, Agarwal, McDaniel, & McDermott, 2011). For example, when learning Swahili-English translations, instead of repeatedly studying word pairs like “kazi - work”, learning would be improved by replacing some of the trials with testing: “kazi - ???”. This general finding of a retrieval practice effect has been consistently described in the literature using different materials such as word lists, text passages, novel facts, and language word pairs.

Although there is currently no agreed-upon mechanism for this effect, it is thought to result from changes to the memory trace associated with practice not present when re-reading. Possible mechanisms include the elaboration of the original memory with additional information (e.g., Carpenter, 2009), increased retrieval cues (e.g., Lehman, Smith, & Karpicke, 2014), increased retrieval strength (e.g., Bjork & Bjork, 1992), or transfer-appropriate processing (e.g., Thomas & McDaniel, 2007). Importantly, prior theories of retrieval practice assume that the underlying learning mechanism applies to all content in the same way. What to do, then, with evidence that although retrieval practice improves learning in some situations, in others, further example study using worked examples is more beneficial (”Worked Example Effect;” Sweller & Cooper, 1985)?

The literature on the worked-example effect has demonstrated that learning benefits from studying examples of how to solve problems instead of practice activities (van Gog, Paas, & van Merriënboer, 2006), or along with practice activities (Renkl, 2005). For example, students learning to calculate the area for the trapezoid would benefit from studying problems where the answer and the steps to solve the problem are worked out, compared to solving the same problem.

Theoretical explanations of the worked-example effect center around two main ideas, potentially complementary: (1) that problem-solving practice puts a greater load on learners’ limited processing capabilities in a manner that is “extraneous” to the learning process (e.g., van Merrienboer & Sweller, 2005) rendering it less effective, and (2) that practice activities are less beneficial because they lack the necessary support to fill-in potential knowledge gaps (e.g., McNamara & Kintsch, 1996). Consistent with these ideas, learning from reading materials can be improved by including pre-questions about relevant parts of the text (e.g., Rickards, Anderson, & McCormick, 1976), or by eliminating unnecessary content and reducing text to its main topics (e.g., Reder, 1980; Reder & Anderson, 1982).

Thus, current evidence suggests that more practice or retrieval can both improve or delay learning and that more study can both improve or delay learning. This apparent contradiction poses both theoretical and practical issues. Theoretically, to which degree do we have a complete understanding of the learning process if opposite mechanisms can yield similar results? Practically, when making suggestions for the application of cognitive science findings to educational contexts, practitioners are left wondering which approach to use and when.

To be clear, ours is not the first attempt at addressing this inconsistency. van Gog & Kester (2012, see also van Gog & Sweller, 2015), proposed that problem complexity was the critical dimension that defined whether retrieval practice or

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

worked examples would improve learning. They proposed that worked examples improve learning of complex problems by reducing cognitive load, whereas practice would improve learning of simpler materials that do not pose the same level of cognitive load. However, as Karpick and Aue (2015, see also Rawson, 2015) pointed out, this explanation does not capture all the evidence. For example, there is ample evidence that retrieval practice improves learning of complex texts (Rawson & Dunlosky, 2011). Ultimately, problem complexity is hard to operationalize, and a lot of the discussion has centered around what constitutes complexity (Karpick & Aue, 2015; Rawson, 2015).

Here we take a different approach. We address this apparent contradiction by empirically testing a possible flexible mechanism that can yield best learning outcomes from retrieval practice in some situations, and from studying examples in others. Our proposal is that retrieval practice improves memory processes and strengths associations, whereas studying examples improves inference processes and information selection for encoding. Thus, when problems include information that must be inferred, combined, or selected from among a complex set of possible pieces of information, studying examples will improve learning. In other situations, retrieval practice will improve learning. This proposal is consistent with previous work showing that retrieval practice improves learning of associations, such as paired associates or text that learners should try to retrieve either verbatim or by putting together several pieces of information (e.g., Karpicke & Blunt, 2011). Conversely, studying examples improves learning of knowledge that requires learners to infer or provide answers to multi-step problems or applying procedures (e.g., learning to calculate the area of a geometric solid, Salden, Koedinger, Renkl, Aleven, & McLaren, 2010).

Importantly, this proposal requires careful identification of which type of knowledge is being used. Associations can be complex and inference-based problems can be simple. Instead, we use the knowledge nomenclature and classification proposed by the Knowledge Learning Instruction framework (KLI; Koedinger, Corbett, & Perfetti, 2012). The KLI framework proposes that learning depends on knowledge and includes a precise classification system for knowledge. Based on analyses from over 360 in vivo studies using different knowledge content, KLI relates knowledge, learning, and instructional events and presents a framework to organize empirical results and make predictions for future research. According to KLI, Instructional Events are activities designed to create learning. Textbooks, lectures and tests/quizzes are examples of commonly used Instructional Events in educational practice. These Instructional Events in turn give rise to Learning Events -- changes in cognitive and brain states associated with Instructional Events. KLI identifies three types of Learning Events: Memory and fluency-building processes, induction and refinement processes, and understanding and sense-making processes. These changes in cognitive and brain states influence and are influenced by the Knowledge Components (KCs) being learned. A KC is a stable unit of cognitive function that is acquired and modifiable. In short, KCs are the pieces of cognition or knowledge and are domain-agnostic. Although Learning Events and Knowledge Components cannot be directly observed, they can be inferred from Assessment Events, or outcome measures, such as exams and discussions.

KLI also offers a taxonomy for KCs based on how they function across Learning Events and relates differences in KCs with differences in Learning Events. In this way, KCs can be classified based on their application and response conditions. Some KCs are applied under unvarying, constant conditions (e.g., paired-associates), while others are applied under variable conditions (e.g., rules). Similarly, the response of the KC can be a single value or constant such as a category label, or it can vary as a function of the variable information extracted in the condition, such as calculating the area of a geometric solid. Thus, according to KLI, facts such as "the capital of France is Paris" are constant application and constant response KCs because there is only one single application of the KC and there is only one response. Moreover, facts require verbatim retrieval and application of studied information (e.g., write the word "Paris" when prompted for the capital of France). Conversely, skills such as equation solving are variable application and variable response KCs because there are multiple different problems that can elicit the same equation solving KC and there are multiple ways to apply this KC across different problems. In this sense, skills require creating generalizations beyond the studied information (e.g., solving an equation that one never saw, using a generalization extracted from studying many examples).

By connecting the type of Learning Event and the associated learning processes with the type of KC being learned, the KLI framework further suggests causal links between instructional principles (e.g., "retrieval practice", "worked-example study"), and changes in learner knowledge. For simple constant KCs such as facts, memory and fluency processes are more relevant. Conversely, for variable KCs such as skills, induction and refinement processes are more relevant. Different learning processes will be optimized by different types of instructional principles, e.g., retrieval practice for facts and studying examples for skills. Thus, different types of KCs will interact with different types of Instructional Principles to create different learning.

Following KLI's framework and nomenclature convention, our proposal is that retrieval practice will be particularly beneficial when learning facts ("Paris is the capital of France"), whereas studying worked examples will be particularly beneficial when learning skills such as equation solving.

In the context of facts ("What is the capital of France?"), learners need to successfully encode all of the information presented and be able to retrieve it later. Learning facts only requires learning the specific pieces of single practice items but does not require any synthesis across practice items. Conversely, in the context of skills ("Calculate the area of a rectangle with the following measurements"), learners need to generalize their knowledge across a series of studied instances. In this sense, learning skills requires identifying which pieces of the information are relevant for encoding and which are not. Put another way, when learning facts all presented information is critical and should be encoded, whereas when learning skills, only a subset of the presented information is relevant to forming an effective generalized skill. Finally, this theoretical proposal is also consistent with procedural differences between research on retrieval practice and example study. Research on retrieval practice generally tests learners' memory on the information presented in repeated trials, whereas research on worked examples generally uses different examples of the same concept in each trial.

To test this hypothesis, in this paper we compare learning outcomes following training of facts and skills, using retrieval practice (Practice-Only) or examples (Study-Practice). For this purpose, we use an equivalent domain and topic (geometry learning), but vary whether learning focuses on fact acquisition (e.g,, "what is the formula to calculate the area of a triangle?"), or skill acquisition (e.g. "what is the area of the triangle below?").

The Experiment

Method

Participants A total of 103 participants volunteered to participate in this study through Mechanical Turk. The whole study took approximately 20 minutes and participants were paid $3.00. No demographic information was collected.

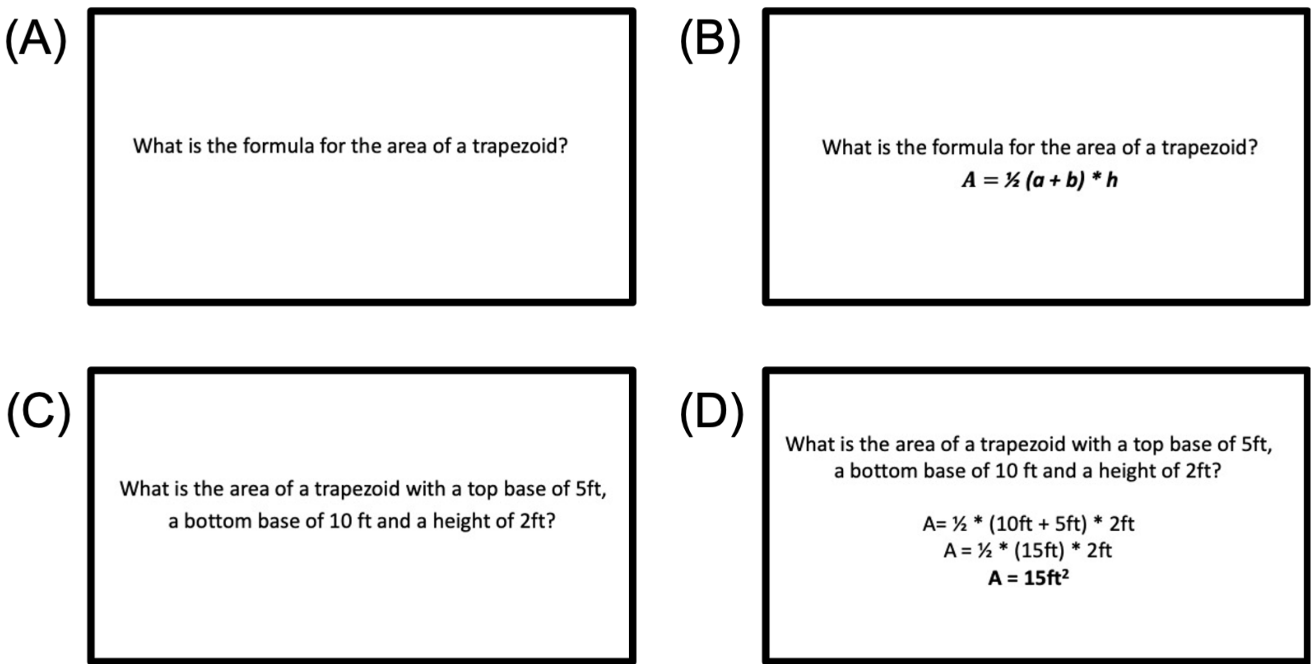

Figure 1: Examples of training trials for trapezoid area. (A) Fact practice trial, (B) Fact-based study trials, (C) Skill practice trial, and (D) Skill study trial.

Participants were randomly assigned to one of four conditions: Practice-Only Training of Facts (N = 32), Study+Practice Training of Facts (N = 23), Practice-Only Training of Skills (N = 20), and Study+Practice of Skills (N = 28).

Data from 8 participants were excluded due to failure to complete the entire experiment (5 from the Practice-only Training of Facts condition, 1 from the Study+Practice Training of Facts condition, and 2 from the Practice-Only Training of Skills condition). The final sample included 95 participants.

Apparatus and stimuli Participants learned how to calculate the area of geometrical shapes (rectangle, triangle, circle, and trapezoid).

We created two multiple-choice tests to be used as pre/posttest. In each test, there were 4 questions about each geometrical shape for a total of 16 questions. For each shape, two questions focused on fact-based knowledge (e.g., "What is the formula to calculate the area of the rectangle?"), and two focused on skill-based knowledge (e.g., "What is the area of a rectangle that is 9 ft wide and 15 ft long?"). Some problems included images and others only text. The problems in the two tests were created to be equivalent.

The training phase involved study of examples and practice memorizing the formulas (Fig 1A) or calculating the area (Fig 1C), depending on the condition (see below for details). In each trial, participants were presented with either a problem to complete or an example to study. For facts the examples were simply the response to the question (see Fig 1A), whereas for skills the examples included the worked out steps to complete answer the problem (see Fig 1D). We created a total of 24 problems (6 per geometrical shape) for each condition. During training, participants were asked to input their answers and no feedback was provided.

To fill the time between training and test, participants completed a series of trivia questions retrieved from the updated and expanded Nelson and Narens (Nelson & Narens, 1980) norms developed by Tauber et al. (2013). Questions of all difficulties were randomly selected to be presented.

Design and Procedure Participants started by completing one of the multiple-choice tests as the pretest. Immediately following the pretest participants started the training phase. During the training phase participants were told that their task was to study the examples and complete the activities in order to learn about the area of these geometrical shapes. For all conditions, problems were presented blocked by geometrical shape, order randomized. The first trial for each geometrical shape was always a study trial in which participants studied an example of a question along with the correct response. In the practice-only conditions participants then completed 3 practice trials for the same geometrical shape before moving to the next geometrical shape. In the study-practice conditions participants then completed a practice test, followed by another study trial, and a final practice test before moving on to the next geometrical shape. When learning facts participants were asked to type the correct formula, when learning skills participants were asked to type the area after calculations. No feedback was presented. Which problems were used was randomized for each participant.

Following the 16 training trials (4 per geometrical shape), participants completed 30 trivia questions randomly selected from a sample of 299 questions. Because we kept the number of questions and not time constant, the duration of this retention interval varied across participants depending on how fast they answered the questions. Immediately following the trivia questions participants completed the other multiple-choice test as the posttest. Which of the two tests was used as the pretest and which was used as the posttest was counterbalanced across participants.

Results and Discussion

Pretest Overall, participants' pretest performance was moderate (M = 0.59 and 0.60, for facts and skills, respectively) and did not differ for facts vs. skills t (188) = 1.26, p = .208.

Pre-Post Change We analyzed posttest performance controlling for pretest performance for each type of trained concept (skills vs. facts), training type (Practice-only vs. Study-Practice), and type of test questions (skills vs. facts). Data were analyzed by fitting a linear mixed-effects model predicting posttest score, using pretest score and duration of retention interval as continuous predictors and type of concept, type of question, and study condition as categorical predictors, as well as their interaction terms. We included

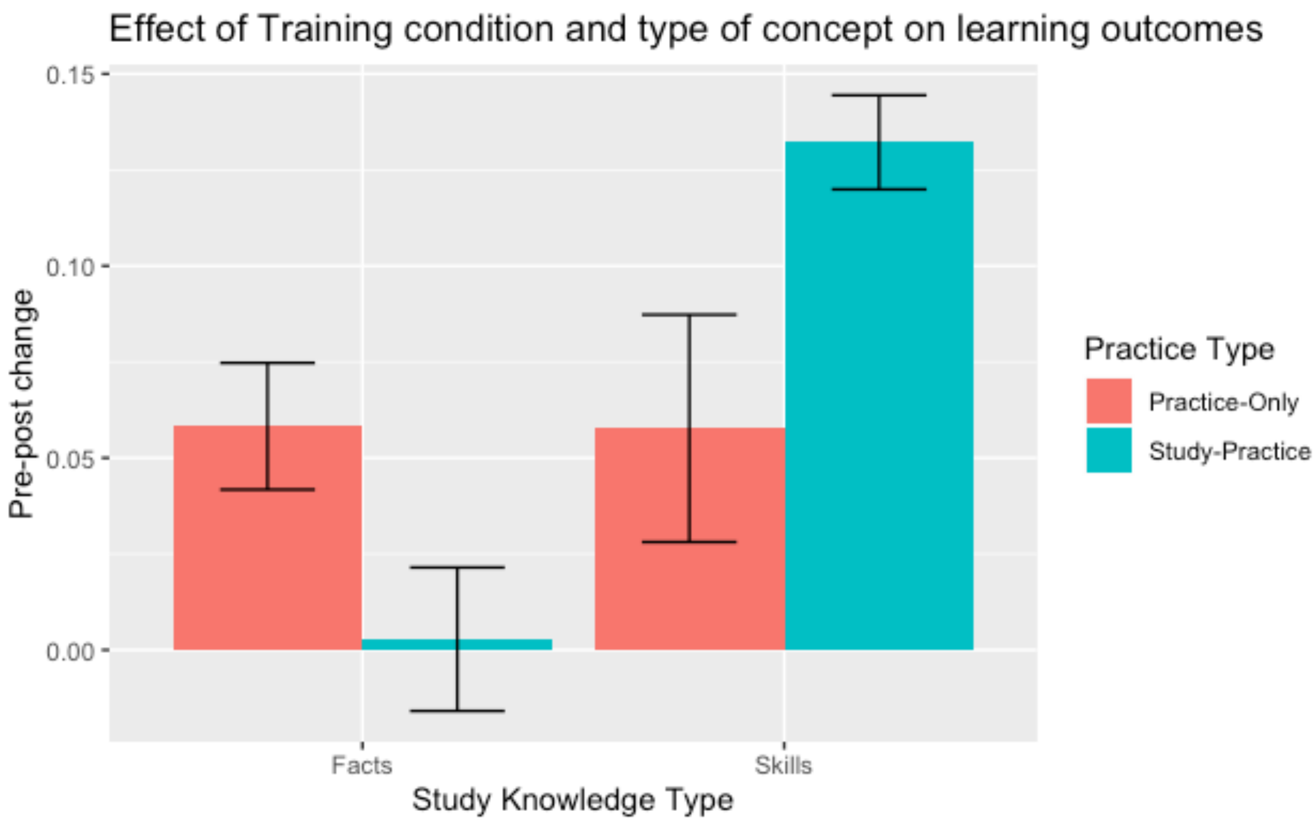

Figure 2: Pre-post change results for different types of concepts and training tasks. Error bars represent the standard error of the mean.

retention interval as a predictor because we did not control its duration and previous research has suggested that the benefit of practice study is moderated by duration of the retention interval (Pan, Gopal, & Rickard, 2016). Although we did not find any difference in pretest scores, including the pretest as a covariate controls for potential differences in previous knowledge between the groups, despite random assignment. For simplicity, we plot pre-post change instead of posttest and pretest values.

As it can be seen in Figure 2, we saw a significant interaction between the type of concept studied and the type of training, $\beta = 0.41$, $t(124.53) = 2.09$, $p = .038$, $d = 0.38$. Thus, whether more practice or study led to better learning depended on the type of concept being learned (skills vs. facts).

However, we found no 3-way interaction with type of test question $\beta = 0.03$, $t(68.73) = 0.133$, $p = .894$, $d = 0.03$, suggesting that this effect is not specific to the type of question being asked and there is some transfer from best learning of skills to fact questions and vice-versa. Interestingly, although fact questions were slightly easier than skill questions $\beta = 0.20$, $t\ (70.33) = 2.05$, $p = .044$, $d = 0.49$, overall performance after learning facts was not different from overall performance after learning skills, $\beta = 0.24$, $t\ (124.86) = 1.88$, $p = .063$, $d = 0.34$.

Finally, contrary to some theoretical predictions, we found no interaction between type of training and retention interval, $\beta = 0.01$, $t(124.52) = 0.834$, $p = .410$, $d = 0.15$, and participants' accuracy responding to fact vs. skill test questions did not vary with retention interval duration, $\beta = 0.02$, $t(70.79) = 1.38$, $p = .173$, $d = 0.33$. Overall, performance following learning facts was slightly worse after longer delays than short delays with no effect of delay for learning skills, $\beta = 0.03$, $t\ (124.74) = 2.20$, $p = .030$, $d = 0.35$. There were no other effects of retention interval or interactions.

Discussion

In this paper we demonstrate that learning is flexible and depending on what is being learned, performance in the same task can vary substantially.

We proposed that, contrary to some theoretical and empirical investigations (Roediger & Karpicke, 2006), learning from practice does not always yield the best outcomes. In fact, learning by alternating study and practice yields better outcomes in some situations. However, our investigation goes beyond this demonstration. We proposed a mechanism through which this flexibility takes place.

We build on the empirical and theoretical understanding proposed by KLI (Koedinger et al., 2012) to identify the specific ways in which the knowledge content changes how the learning processes involved in the effect of testing practice operate. The general proposal is that increased information presented during encoding requires increased selectivity (identifying relevant elements for encoding) for successful induction and refinement. Increased selectivity in turn requires encoding processes that successfully direct the learner (either intrinsically or extrinsically) towards the relevant information. Studying examples -- as opposed to further retrieval practice -- can play a key role in this aspect. Mechanistically, our proposal is that studying examples guides attention and selectivity towards a subset of the presented information. Subsequent retrieval practice will improve memory and consolidation of this selected information. In this context, when learning facts, no selectivity is required and thus studying examples will not contribute to better learning outcomes, and might even delay it. Conversely, when learning skills, retrieval practice without substantial time dedicated to studying examples could lead to strengthening encoding and consolidation of irrelevant information that does not allow for successful induction and generalization, thus resulting in worse learning outcomes.

In sum, the hypothesis put forward here is that studying examples changes the learning process by introducing selectivity about what is relevant and should be encoded. This change, however, is only going to be beneficial if the learning context requires it. That is, increased selectivity of encoding introduced with further study of worked examples will not positively influence learning when the information presented is reduced or no generalization is required. Finally, this proposal is also consistent with procedural differences between research on retrieval practice and example study; whereas research on retrieval practice has used mostly repeated information across trials and tested learners' memory for that information, the research on worked examples has used mostly different examples of the same concept in each trial.

Conclusion

A distinctive characteristic of human learning is our capability to flexibly acquire a wide range of rich and complex forms of knowledge (e.g., first and second languages, chess and golf, math and science, collaboration and learning strategies, etc.) and get better at acquiring new knowledge as we accumulate knowledge (e.g., learning physics is much easier after having learned how to read and do algebra). What flexible learning mechanism makes human learning this smartly nuanced? Here we started to approach this question by investigating how the same mechanism can improve learning outcomes in one situation but not in others.

Finally, by highlighting the adaptive nature of learning, we hope to not only address outstanding apparent inconsistencies in the literature, but also provide a mechanistic view of learning across multiple situations beyond describing what works (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Only by moving beyond demonstrations of what works, towards demonstrations of what works when along with the precise mechanisms of learning yielding such interactions will we be able to understand human learning and improve it where needed.

Acknowledgments

This work was supported by a National Science Foundation grant (BCS-1824257). All data presented here are available

from LearnShepere.org's DataShop: https://pslcdatashop.web.cmu.edu/Files?datasetId=5189.

References

Bjork, R. A., & Bjork, E. L. (1992). *A new theory of disuse and an old theory of stimulus fluctuation.* In From learning processes to cognitive processes: Essays in honor of William K Estes (Vol. 2, pp. 35–67). Hillsdale, NJ: Erlbaum.

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. doi: 10.1037/a0017021

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques. *Psychological Science in the Public Interest*, *14*(1), 4–58. doi: 10.1177/1529100612453266

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. Science, 331(6018), 772–775.

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review, 27*(2), 317-326.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive science, 36*(5), 757–798.

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an Episodic Context Account of Retrieval-Based Learning: Dissociating Retrieval Practice and Elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(6), 1787–1794.

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes, 22*(3), 247–288.

Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior, 19*(3), 338–368.

Pan, S. C., Gopal, A., & Rickard, T. C. (2016). Testing with feedback yields potent, but piecewise, learning of history and biology facts. *Journal of Educational Psychology, 108*(4), 563–575.

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied, 14*(2), 101–117

Rawson, K. A. (2015). The status of the testing effect for complex materials: Still a winner. *Educational Psychology Review*, *27*(2), 327-331.

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

Reder, L. M. (1980). A comparison of text and their summaries: memory consequences. *Journal of Verbal Learning and Verbal Behavior, 134*, 121–134.

Reder, L. M., & Anderson, J. R. (1982). Effects of spacing and embellishment on memory for the main points of a text. *Memory & Cognition, 10*(2), 97–102.

Renkl, A. (2005). *The worked-out-example principle in multimedia learning.* In R. Mayer (Ed.), The Cambridge Handbook of Multimedia Learning. Cambridge University Press.

Rickards, J. P., Anderson, M. C., & McCormick, C. B.
(1976). Processing effects of common-word and number questions inserted in reading materials. *The Journal of Educational Research, 69*(7), 274–277.

Roediger, 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*(4), 382–395.

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

Salden, R. J. C. M., Koedinger, K. R., Renkl, A., Aleven, V., & McLaren, B. M. (2010). Accounting for Beneficial Effects of Worked Examples in Tutored Problem Solving. Educational *Psychology Review, 22*(4), 379–392.

Sweller, J., & Cooper, G. A. (1985). The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra. *Cognition and Instruction, 2*(1), 59–89.

Tauber, S. K., Dunlosky, J., Rawson, K. A., Rhodes, M. G., & Sitzman, D. M. (2013, December). General knowledge norms: Updated and expanded from the Nelson and Narens (1980) norms. *Behavior Research Methods, 45*(4), 1115–1143.

Thomas, A. K., & McDaniel, M. A. (2007). The negative cascade of incongruent generative study-test processing in memory and metacomprehension. *Memory & Cognition, 35*(4), 668–678.

van Gog, T., Paas, F., & van Merrie ̈nboer, J. J. G. (2006). Effects of process-oriented worked examples on. *Learning and Instruction, 16*, 154–164.

van Gog, T., & Kester, L. (2012). A test of the testing effect: acquiring problem-solving skills from worked examples. *Cognitive Science, 36*(8), 1532-1541.

van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: The testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review, 27*(2), 247-264.

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review, 17*(2), 147–177