UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Not all Active Learning is Equal: Predicting and Explaining Improves Transfer Relative to Answering Practice Questions

Permalink

Journal

Authors

Carvalho, Paulo F
Manke, Kody J
Koedinger, Kenneth R

Publication Date

2018

Not all Active Learning is Equal: Predicting and Explaining Improves Transfer Relative to Answering Practice Questions

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

Kody J. Manke (kmanke@andrew.cmu.edu)
Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213 USA

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

Abstract

We compared students' exam performance following one of two different types of active learning assignments. In one version students read text describing experimental evidence for the principle being studied. In the other version, students instead created a hypothesis and explanation, and then studied and explained the results. The content was matched across conditions. Students performed better in exams requiring generalization to novel situations, after providing hypotheses and explanations than after reading the text and answering questions about it. These results suggest that prediction and explanation cycles might be a better active learning approach to promote generalization and transfer than practice questions.

Keywords: predict-observe-explain; active learning; retrieval practice

Introduction

Students learn better when they engage in active learning (Freeman et al., 2014; Wieman, 2014). Yet, much instructional practice emphasizes passive learning such as reading text, attending lectures, and watching videos. Contrary to evidence of the clear benefits of active learning, students (and a surprisingly high number of instructors) feel that passive strategies such as re-reading are useful study methods (Morehead et al., 2016).

This disconnect between evidence and practice highlights the need to develop active learning practices that are grounded in empirical evidence and can support effective learning. One important step is understanding which types of active learning practices support transfer of different types of knowledge. For example, an active learning practice that supports memorization processes might not be equally effective for generalization (Koedinger et al., 2011).

As an initial step towards developing better active learning practices, in this paper we compare two different types of active learning activities in a psychology course and evaluate their impact on students' learning outcomes. Specifically, we compare active learning that involves responding to questions after reading a text to active learning that involves generating hypotheses and explanations.

Previous research has shown that, even following an introductory course in psychology, students often lack the ability to provide clear scientific reasoning and identify methodological issues (Lawson et al., 2015). These difficulties include generating testable hypotheses and connecting scientific evidence to theory. One possible explanation is that the learning practices used do not match the learning goals. Despite the expectation, across most STEM courses, that students learn scientific principles through the study of empirical evidence, most learning practices are either passive (e.g., reading) or focus on active learning that promotes memorization of evidence, instead of extrapolation (e.g., retrieval practice or pre-questions).

In this research, we compare the use of practice questions and another active learning practice: the predict-observe-explain process. As we discuss in greater detail below, the use of practice questions has been shown to successfully improve learning and memory for studied facts, whereas the predict-observe-explain process promotes predicting and explaining evidence and therefore has the potential to foster scientific understanding and inductive processes.

Practice questions approach

Learning from reading is remarkably poor compared to active learning. For example, completing more practice activities is a better predictor of learning in online courses than completing more readings (Carvalho et al., 2017). Furthermore, practices that reduce the amount of reading often improve learning; across multiple studies, learners remembered more facts when they read only the summaries instead of the textbook prose, even when the critical information was highlighted (Reder & Anderson, 1980a).

Several active learning strategies have been suggested to improve learning from passive information. These often include questions either before the text (pre-questions; e.g., Rickards, 1976a), along with reading the text (Rickards, 1976b), or after reading the text instead of re-reading (Roediger & Karpicke, 2006). The inclusion of practice questions, in any of these ways, has been shown to improve learning compared to having no questions (Pressley, Tanenbaum, McDaniel, & Wood, 1990; Richland, Kornell, & Kao, 2009; Rickards, 1976b, 1976a), re-reading the materials instead of answering the questions (Butler & Roediger, 2007; Richland et al., 2009), or reading the questions along with the answers (Carpenter et al., 2017).

Several (not mutually exclusive) mechanisms have been suggested for why practice questions improve learning. Practice questions orient the learner towards the critical aspects of the text or video (Reder, 1980b), they require the learner to retrieve previous information (Roediger &

Karpicke, 2006), and they act as metacognitive checks on the learners' knowledge (Bjork et al., 2013). Although the exact mechanisms of why practice questions improve learning are still debated, the active learning nature of practice questions compared to only reading is undeniable.

Importantly, most of the research on the benefits of practice questions has focused on retrieval of information. That is, the amount of information learners can successfully recall from the text or video provided. Thus, it is possible that, despite its high effectiveness in improving learners encoding and recall of information (compared to passive learning), practice questions might not elicit induction and refinement processes (Koedinger et al., 2011). When the goal of learning is to generalize or induce scientific principles from specific empirical examples, practice questions may be insufficient.

The Predict-observe-explain approach

Another active learning approach is the use of the predict-observe-explain process to guide learners through learning materials, often regarding empirical results (White & Frederiksen, 1998). In the predict-observe-explain (POE) process, learners are introduced to a scientific question (e.g., *Are we influenced by group pressure?*) and an experimental setup (e.g., *Two groups are answering questions about line lengths. In one group participants answer questions individually, in the other group participants answer questions in a group in which 4 people previously provided the ostensive wrong answer*). Learners are then asked to make a prediction about the study results (*Predict* step). After learners make their predictions, they are presented with the experimental results (*Observe* step), usually in the form of a graph or table, and then asked to explain why the experimental manipulation yielded such results (*Explain* step).

When compared to business-as-usual classroom practices (passive learning), POE has been shown to improve learning (Karamustafaoglu & Mamlok-Naaman, 2015; Kibirige, Osodo, & Tlala, 2014; White & Frederiksen, 1998). The exact mechanisms through which POE improves learning are not well understood. It may involve the ability to directly address *a priori* misconceptions students might have with evidence that directly contradicts it (Kowalski & Taylor, 2009). It may also involve creating explanations, which might improve learning as repeatedly shown with the self-explanation effect (VanLehn, Jones, & Chi, 1992). In particular, creating explanations has been shown to promote generalization and transfer (Lombrozo, 2006), albeit hurting memory (Williams & Lombrozo, 2010).

Regardless of the exact mechanism, POE has the distinctive feature of providing active learning, similar to practice questions. However, POE emphasizes generalization and extrapolation from data, whereas practice questions emphasize rote memorization of the information provided in the text. Thus, it is possible, even likely, that POE leads to better learning than practice questions when we test students' ability to generalize evidence covered in class.

The present study

The main goal of this initial study is to compare the learning benefits of two active learning approaches – practice questions and predict-observe-explain – for generalization of empirical evidence to new situations. Unlike previous research, we compare a POE approach not with business-as-usual passive learning practices, but with another active learning practice shown to also improve learning.

We used an in-vivo procedure in which we embedded an experimental manipulation in the instructional activities of an introductory undergraduate Social Psychology course. To increase power and decrease potential issues related to individual differences, we used a within-subjects design. There were two conditions: a practice questions condition and a predict-explain-observe-explain (PEOE) condition. In the practice questions condition, students read the description of three social psychology studies for two different social psychology topics and were asked to answer questions about those materials. In the PEOE condition, students studied the same empirical studies but were only provided the research question and experimental design before being asked to predict the outcome, explain their prediction, observe a plot of the results, and explain the results observed. We included the additional explanation step compared to the typical POE to maximize the potential benefit of self-explanations associated with POE.

Students completed a series of open-ended and multiple-choice questions as part of their exam. None of the questions in the exam probed memory for the specific studies practiced or asked about specifics of those studies. Instead, questions asked students to apply general knowledge about the principles to novel situations or relate it to other principles covered in class. We measured students' performance on exam questions about the topics covered in the activity, general exam performance, time spent completing the activities, and performance on the activities themselves.

We predicted that active learning that involves PEOE will result in improved generalizable learning and transfer to other topics.

Method

Participants

One hundred nineteen undergraduate students volunteered to participate in the study as part of their Social Psychology course taught by the second author at Carnegie Mellon University. All students completed all conditions (see below for details), order counterbalanced across participants. Forty-four students did not complete at least one assignment before each exam and their data were excluded from analyses. The final sample includes data from 75 students.

Materials

Study materials were created using the Open Learning Initiative (OLI) authoring tools and distributed to students as assignments using the course's Canvas website.

Table 1: Topics of the six assignments and schedule.

Assignment	Topic	Course week
1	Obedience	1
2	Cognitive Dissonance	2
	Exam 1	5
3	Conformity	8
4	Social Facilitation	9
	Exam 2	10
5	Stereotype Threat	11
6	Misattribution of Arousal	14
	Exam 3	15
	Final Exam	Finals week

For assignments 3-6 (see Table 1), we created two versions: Practice Questions and PEOE. The first two assignments did not have a PEOE version and were used as a baseline condition.

Figure 1: Example of one of the assignments in the Practice Questions condition.

Each assignment covered three empirical studies in Social Psychology about a topic covered in that unit. The studies were chosen for their relevance and representativeness for the topic but were not covered by the instructor in class (though the topic was). Table 1 includes all topics covered.

The Practice Questions version of the materials included a short description of the study procedure and background, and description of the hypothesis, predictions, results (including a plot), and conclusions of the study (see Figure 1). After the text, students were asked to answer four open-ended questions about the text they just read: (1) *What were the researchers' predictions?*, (2) *Why did the researchers make those predictions?*, (3) *What did the results show?*, and (4) *Why did participants respond that way?*. Answers to each of these questions were clearly and succinctly stated in the text students had just read. No feedback was provided for any of the questions. Although questions were presented after the text, because they were in the same page, students were free

to use the questions as pre-questions or along with the text as they wished.

The PEOE version of the materials started with a brief description of the study procedure and background (same paragraph as in the Practice Questions version). After that paragraph a series of questions instantiating the predict-explain-observe-explain procedure were presented, one at time: (1) *What do you expect will happen?*, (2) *Why do you think [hypothesis selected]*?, (3) *The graph above shows the results of this study. What do the results show?*, and (4) *Why do you think the results show [results]?* (see Figure 2 for an example)

Figure 2: Example of one of the assignments in the PEOE condition.

Questions (1) and (3) were multiple-choice questions in which the same three possible hypotheses were presented, but the graph with the experimental results was presented only along with question (3) and not before students answered (1) and (2). Questions (2) and (4) were open-ended questions. Feedback was provided only to questions (3) and (4). Feedback for question (3) indicated the correct description of the results and feedback for question (4) indicated the conclusions extracted by the experimenters in the target-study and matched the corresponding paragraph in the Practice Questions version of the materials.

The two versions of the materials were closely matched for active engagement/retrieval (in all versions students had to answer questions) and content (the questions and text were matched across versions and the same text was used across conditions where needed).

Students completed three non-cumulative exams during the semester and a cumulative final exam. For each exam, we created five questions targeting the topic covered in the activities. Two were open-ended questions asking students to apply the principles to novel situations, two were multiple-choice questions targeting students' understanding of the principle, and another was a multiple-choice question about general understanding of research methods. None of the exams included questions about the specific studies covered

in the activities but included other questions relating activity topics with other class topics. The open-ended questions were scored by a team of course TAs, blind to condition assignment

Design and Procedure

The study took place over a semester and the study activities were assigned as homework. Students completed two assignments before each exam. The first two assignments had only a Practice Questions version and were used as a baseline condition for all students. A crossover design was used such that students were randomly assigned to one of two groups. The first group completed the Practice Questions version of assignments 3 and 4 (before Exam 2) and the PEOE version of assignments 5 and 6 (before Exam 3). The other group completed the inverse versions – the PEOE version of assignments 3 and 4, and the Practice Questions version of assignments 5 and 6. All assignments were due the week the topic was covered in class, and students received course points if they completed the assignment, regardless of performance on the assignment. All students were required to complete all course exams.

Results

We normalized all measures by calculating z-scores of all predictor and outcome measures. This allowed us to compare estimates across analyses as normalized effect sizes and addressed possible scaling issues related to using very different variables in regression models.

Time spent completing the activities

Students spent on average more time completing the Practice Questions version ($M = 53$ minutes, $SD = 54$ minutes) than the PEOE version of the assignments ($M = 31$ minutes, $SD = 35$ minutes). This difference was statistically significant when controlling for average time spent on baseline assignments ($M = 55$, $SD = 53$) and counterbalancing condition, $\beta = -0.48$, $t (74) = -4.14$, $p < .00001$.

We also analyzed the logged data to ascertain that students were completing the practice questions in the corresponding version of the assignments, and whether they were doing so before or after reading the text. Only one student did not complete all questions for all assignments (one question was left blank). The average number of words written in the answer to the questions was 26 ($SD = 7$), and initial inspection of the responses suggests that students actively tried to respond and not only directly copy from the text or provide random strings. Finally, on average, students spent less than a minute on the page before clicking on one of the question activities, suggesting that students were using the questions as guides to read the text ($M = 4.81$ seconds, $SD = 5.01$ seconds), and not the other way around.

Exam grades

Next, we looked at students' exam performance across conditions, controlling for time spent on the assignments, counterbalancing condition, performance on the baseline exam, and time spent on the baseline assignments. Students performed better on the exams covering topics practiced using the PEOE version of the assignments than in those with topics practiced using the Practice Questions version of the assignments (see Figure 3), $\beta = 0.13$, $t (82.18) = 2.75$, $p = .007$. The effect is even larger when we do not control for time spent in the activities, $\beta = 0.31$, $t (74) = 2.65$, $p = .010$.
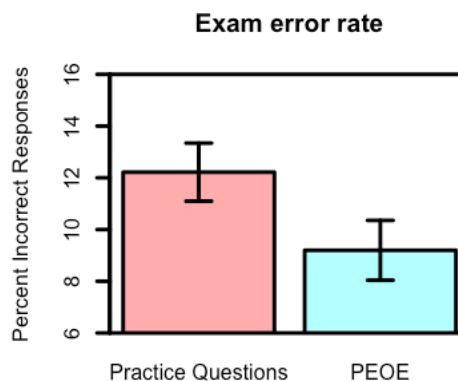

Figure 3. Students' performance on the exams following Practice Questions and PEOE activities.

We also looked at the effect of the version of the assignments (PEOE vs. Practice Questions) on the questions we included in the exam specifically only about the studies covered in the assignments. Students showed only slightly fewer errors on exam questions specifically about the topics covered in the PEOE version of the assignments ($M = 0.07$, $SD = 0.10$), compared to the Practice Questions version of the assignments ($M = 0.08$, $SD = 0.10$). This difference was not statistically significant when controlling for time spent in the activities, counterbalancing condition, performance on the baseline exam, and time spent on the baseline assignments, $\beta = 0.01$, $t (74) = 0.07$, $p = .945$. One potential reason for this result is the compressed scale. We included only four questions referring only to the specific topics covered in the activities and, overall, the range was smaller (between 0%-45% error rate), compared to the range of results for the entire sample (between 0%-58% error rate). Importantly, performance on the assignment-specific questions was a strong predictor of performance on the exam, $\beta = 0.50$, $t (71) = 8.07$, $p < .00001$. This suggests that the overall exam differences may be related to potential spillover from stronger understanding of a considerable portion of the materials covered on the exam, even if the remaining questions did not uniquely target assignment topics.

Quality of students' predictions

Beyond time spent on the assignments (included in all analyses above), does accuracy on the assignment itself predict exam performance? Specifically, do students who make correct predictions in the first step of the PEOE version benefit more from that assignment activity compared to those who make the wrong prediction? To answer this question, we compared performance on the exam for students in the PEOE condition as a function of their accuracy in the predict step.

On average, students made the correct prediction on the first try 54% of the time (*SD* = 18), and 53% of the time when considering all tries (SD = 18; mean number of tries = 1.32). Importantly, students who made correct predictions on the first try performed better in the exam compared to those who did not, β = 0.21, t (71) = 2.17, p = .033.

For comparison, and to determine if this effect is unique to the *predict* step or reflects general ability or compliance differences among students, we compared student exam performance depending on their accuracy responding to the question about the plot – the *observe* step. The answer to this question was presented in the plot if students successfully read it. Students provided the correct response to this question on the first try 81% of the times (SD = 15), and 82% of the times when considering all attempts (SD = 14, mean number of tries = 1.28). We found no difference on exam performance depending on the students' accuracy describing the graph, β = 0.140, t (71) = 1.44, p = .153. Finally, even when controlling for accuracy on the predict step, overall students performed better in the exam following PEOE activities compared to Practice Questions, β = 0.34, t (82) = 2.78, p = .007.

Discussion

The results of this study suggest that active learning activities using a predict-explain-observe-explain approach take less time to complete than reading along with practice questions while simultaneously yielding better learning, as measured by exam performance.

The predict-observe-explain approach had been used before with positive learning outcomes. However, previous studies focused on comparing the POE approach to business-as-usual classroom activities. The present study is, to the best of our knowledge, the first study that demonstrates the benefits of PEOE when compared to another active learning activity recognized to improve learning.

Reading text materials and answering practice questions is an intensive, time-consuming activity and students in the current study seem to use the practice questions as reading guides, which has been shown to improve learning (Carpenter et al., 2017). The effortful use of practice questions could make this type of activity a "desirable difficulty" (Bjork, 1994) – a group of practices that, despite involving greater effort and worse immediate outcomes, improve learning in the long run. However, the current study shows that, despite high compliance and student engagement in the practice question assignments, students performed better on topics practiced using the PEOE approach.

Previous research has shown that practice questions are effective at improving learning compared to many other approaches (Roediger & Karpicke, 2006). Critically, the benefit of PEOE compared to practice questions seen here is likely to be tied to the different types of assessments and corresponding types of learning (Koedinger et al., 2011). Previous work assessing the benefit of practice questions has focused mostly on verbatim memory for the materials studied or conceptual extrapolations that had been probed in the practice questions (Butler, 2010; Roediger & Karpicke, 2006). In the present work, however, the exam asked students to generalize psychological principles from the materials studied and connect them to other concepts. Under these circumstances, learners who created and explained hypotheses performed better on the exam than those who read the same information and answered questions about it. Although we did not include on the exam questions testing verbatim memory for the activities in this first study, future work will further test this hypothesis by directly comparing performance on verbatim and generalization questions. This hypothesis is also consistent with previous evidence showing that learners who provided explanations in a categorization task showed better category generalization but worse memory for specific items studied compared to learners asked to describe the items (Williams & Lombrozo, 2010).

Any of the differences between PEOE and practice questions could account for the results presented here, including the reduced text, asking for explanations, or step-by-step presentation. Although the main goal of this study was not to identify the specific mechanisms that make POE (and its PEOE counterpart used here) good for learning, there are several suggestions from the current literature. Previous work has suggested that addressing previous misconceptions by providing hypotheses before seeing the results might improve learning of scientific phenomena. Our results do not seem to support this hypothesis. Students who provided the correct hypothesis on the first try performed better in the exam than those who did not.

An important aspect of PEOE is the several steps of explanation required. Providing explanations – even if incorrect – might improve learning and generalization by promoting integration of new information with previous knowledge (Chi, 2000), and by guiding learners' attention towards structural features that are relevant for generalization (Lombrozo, 2006; Rittle-Johnson, 2006). Thus, it is possible that, compared to practice questions, PEOE led learners to connect the information in the assignments with other materials from the textbook or class and emphasized critical similarities and differences among several examples, within the assignments and across sources.

Importantly, the present study used an in-vivo approach to experimentation. That is, we implemented the manipulation as part of regular classroom activities using relevant materials. This approach increases the external validity of the results and their direct relevance for both theory and practice, while maintaining precise control over the manipulation.

In sum, the present results suggest that asking students to predict and explain information yields better generalization and transfer than other practices known to improve learning over passive learning. Future research is necessary to test the generalizability of the present results (e.g., does PEOE work with other materials beyond scientific results?) and to probe the exact mechanisms of action. The present results also reiterate the importance of considering the whole learning context when identifying the best learning approach (Carvalho & Goldstone, 2015; Koedinger et al., 2011).

Acknowledgements

References

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-Regulated Learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology*, *64*(1), 417–444.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1118–1133.

Butler, A. C., & Roediger, 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., Rahman, S., Lund, T. J. S., Armstrong, P. I., Lamm, M. H., Reason, R. D., & Coffman, C. R. (2017). Students' Use of Optional Online Reviews and Its Relationship to Summative Assessment Outcomes in Introductory Biology. *Cell Biology Education*, *16*(2), ar23.

Carvalho, P. F., & Goldstone, R. L. (2015). What you learn is more than what you see: What can sequencing effects tell us about inductive category learning? *Frontiers in Psychology*, *6*.

Carvalho, P. F., McLaughlin, E. A., & Koedinger, K. R. (2017). Is there an explicit learning bias? Students beliefs, behaviors and learning outcomes. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), P*roceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 204–209).

Chi, M. T. H. (2000). *Self-explaining: The dual processes of generating inference and repairing mental models. In R. Glaser (Ed.), Advances in instructional psychology: Educational design and cognitive science. Vol. 5 (pp. 161-238)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *PNAS*, *111*(23),8410–8415.

Karamustafaoglu, S., & Mamlok-Naaman, W. (2015). Understanding Electrochemistry Concepts using the Predict-Observe-Explain Strategy. *Eurasia Journal of Mathematics, Science & Technology Education*, *11*(5),

Kibirige, I., Osodo, J., & Tlala, K. M. (2014). The Effect of Predict-Observe-Explain Strategy on Learners' Misconceptions about Dissolved Salts. *Mediterranean Journal of Social Sciences*.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2011). The Knowledge-Learning-Instruction (KLI) Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive Science*, *36*(5), 757–798.

Kowalski, P., & Taylor, A. K. (2009). The Effect of Refuting Misconceptions in the Introductory Psychology Class. *Teaching of Psychology*, *36*(3), 153–159.

Lawson, T. J., Jordan-Fleming, M. K., & Bodle, J. H. (2015). Measuring Psychological Critical Thinking: An Update. *Teaching of Psychology*, *42*(3), 248–253.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–470.

Morehead, K., Rhodes, M. G., & DeLozier, S. (2016). Instructor and student knowledge of study strategies. *Memory*, *24*(2), 257–271.

Pressley, M., Tanenbaum, R., McDaniel, M. A., & Wood, E. (1990). What happens when university students try to answer prequestions that accompany textbook material? *Contemporary Educational Psychology*, *15*(1), 27–35.

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

Reder, L. M. (1980b). The role of elaboration in the comprehension and retention of prose: A critical review. *Review of Educational Research*, *50*(1), 5–53.

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

Rickards, J. P. (1976a). Interaction of position and conceptual level of adjunct questions on immediate and delayed retention of text. *Journal of Educational Psychology*, *68*(2), 210.

Rickards, J. P. (1976b). Type of verbatim question interspersed in text: A new look at the position effect. *Journal of Reading Behavior*, *8*(1), 37–45.

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, *77*(1), 1–15.

Roediger, H. L., & Karpicke, J. D. (2006). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, *1*(3), 181–210.

VanLehn, K., Jones, R. M., & Chi, M. T. (1992). A model of the self-explanation effect. *The Journal of the Learning Sciences*, *2*(1), 1–59.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students. *Cognition and Instruction*, *16*(1), 3–118.

Wieman, C. E. (2014). Large-scale comparison of science teaching methods sends clear message. *PNAS 111*(23), 8319–8320.

Williams, J. J., & Lombrozo, T. (2010). The Role of Explanation in Discovery and Generalization: Evidence From Category Learning. *Cognitive Science, 34*(5), 776–806.