Guerrero, T. A. & Wiley, J. (2018). Effects of text availability and reasoning processes on test performance. In *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 1745-1750). Madison, WI: Cognitive Science Society. https://cogsci.mindmodeling.org/2018/papers/0336/0336.pdf

Effects of text availability and reasoning processes on test performance

Tricia A. Guerrero (tguerr9@uic.edu)
Department of Psychology, 1007 W. Harrison Street
Chicago, IL 60607 USA

Jennifer Wiley (jwiley@uic.edu)
Department of Psychology, 1007 W. Harrison Street
Chicago, IL 60607 USA

Abstract

Learning from expository science texts is challenging. These studies explore whether difficulties can be attributed to poor memory or poor reasoning. To eliminate the need for memory during testing, some students took the tests with the texts available. To test for the effects of reasoning on performance, some students were prompted to engage in explanation activities during or after reading. The effects of these manipulations were tested on text-based and inference questions. Allowing the reader access to the texts during testing improved performance for text-based questions. In contrast, engaging in explanation activities during reading improved performance on inference questions. These results suggest that achieving a better understanding from expository texts depends on engaging in constructive reasoning processes, and not simply improving memory for the texts.

Keywords: Text comprehension; Explanation; Inferences; Situation model; Learning from text

Introduction

The goal of reading an expository science text is often for the student to construct a situation model (Kintsch, 1994) or a mental model of a scientific phenomenon, system or process (Graesser & Bertus, 1998; Mayer, 1989). It is the development of a situation or mental model that represents the understanding of how or why a phenomenon occurs, and this understanding is what allows the reader to transfer their knowledge to new contexts (Mayer, 1989). However, most research on text comprehension shows that students struggle with learning from expository science texts. Even college-aged students are notoriously poor at learning from expository science texts, despite the fact that much instruction still involves self-regulated study from their textbooks. In two experiments, the present line of research explored two possible sources of difficulty when readers are tasked with learning from expository science text: poor performance due to poor memory for the text in Experiment 1, or due to a failure to engage in appropriate reasoning processes in Experiment 2.

Experiment 1

The first experiment tested the possibility that one reason why readers may show poor performance is due to poor memory for the information that they read. A common approach that has been used to test whether memory for the text is an obstacle is by removing this source of difficultly. This has been accomplished by providing the reader access to the text while they are testing; hence, the reader no longer has to rely on memory to answer the test questions. If poor memory for the text is one reason why readers are struggling, then when that difficulty is removed, it would be expected that performance should improve. In one study using a text-availability manipulation, Ozuru, Best, Bell, Witherspoon, and McNamara (2007) had undergraduate participants read an expository science text written at a ninth grade level. Participants answered test questions either without the text or with the text as a reference. When participants had access to the text during the testing period, performance on both text-based and inference questions was improved. In contrast, Ferrer, Vidal-Abarca, Serrano, and Gilabert (2017) had middle-school students read a single expository text written at a grade-appropriate reading level. Text availability was manipulated as a between-participants factor. An interaction between text availability and question type indicated that when participants had access to the text during the test, performance was improved for text-based questions only. However, no differences were observed for inference questions. A key difference between these studies may have been the difficulty of the texts and whether readers were given readings below or at their grade level. Based on both these results, it was predicted that having the text available would improve performance on text-based questions. In addition, because the present experiment used expository texts that were written at an appropriate grade level for undergraduates, it was predicted that access to the text might be less likely to have an effect on inference-based questions.

Method

Participants Participants (60 females; $M_{\text{age}} = 18.4$, $SD_{\text{age}} = .83$) were 96 undergraduates who received course credit for their participation in the experiment through the introductory psychology subject pool. Participants were randomly assigned to one of two conditions: having access the text while testing (with-text) or testing without the text available (without-text). A between-participants design was selected specifically to eliminate the possibility of carryover effects within participants. Each participant read and was tested on two of the six texts. Texts were fully counterbalanced so that each of the 6 texts were assigned to 16 participants in each of the 2 conditions, resulting in 192 observations.

Materials Passages used in this experiment (adapted from Thiede, Wiley, & Griffin, 2011) introduced six different phenomena (e.g. how volcanic eruptions occur, how food allergies develop, why ice ages occur, what causes the differences in scores on IQ tests, how monetary policy affects the economy, how evolution occurs). (See Table 1 for an example text excerpt and example questions.) The texts were between 650 and 1000 words in length, and were written at the 11-12th grade level with reading ease scores in the difficult range of 31-49 according to Flesch-Kincaid.

The test booklet contained ten questions for each topic, five of which were text-based questions and five inference questions (also based in Thiede, Wiley, & Griffin, 2011). Text-based questions were either explicitly mentioned in the text or could be found through a verbatim or paraphrased lexical search. Inference questions required the reader to apply the information presented in the text to a new situation or arrive at an answer by integrating multiple pieces of information from across parts of the text. On the tests, inference questions were presented first, followed by the remaining five text-based questions. The answers were presented in multiple-choice format with the correct answer and three distractor options. Distractor options were similar to the correct answers and contained words from the texts.

Procedure Participants were randomly assigned to read two of the six topics. They were first given an opportunity to read through both texts at their own pace. Following the reading phase, participants were presented with the final tests presented one at a time in the same topic order as they were read. In the without-text condition, participants took the test without access to the text. In the with-text condition, participants took the test with access to the text and were encouraged to use the text while answering the test questions.

Results

As shown in Figure 1, a 2 (Text Availability: With, Without) x 2 (Question Type: Text-based, Inference) repeated measures analysis of variance (ANOVA) indicated a significant interaction, $F(1,190) = 28.83$, $p < .001$, $\eta^2 = .06$. There was also a main effect of condition, $F(1,190) = 11.77$, $p < .001$, $\eta^2 = .04$, and a main effect of question type, $F(1,190) = 136.83$, $p < .001$, $\eta^2 = .23$ that were subsumed by this interaction. Overall, inference questions were more difficult than text-based questions, and test performance was better in the text-available condition. However, the interaction emerged because readers who had the text available outperformed those who did not have the text available during testing on text-based questions, $t(190) = 6.5$, $p < .001$, $d = .94$, but not on inference questions, $t < 1$.

Table 1.
Text Excerpt and Example Questions

Text: Why do ice ages occur?	
The more CO2 there is in the atmosphere, the more long-wave radiation is kept from leaving the Earth. The more radiation that is trapped, the hotter the Earth becomes. This trapping of radiation works like a gardener's greenhouse, and this phenomenon is commonly known as the 'Greenhouse Effect'. When a region receives less solar radiation, there is less energy to warm that area. Less heat energy leads to cooler temperatures. Cooler temperatures can cause more snow and ice to form. Snow and ice on mountaintops can reflect what little solar energy reaches the surface of the Earth back into space. The formation of snow and ice can also steal large amounts of CO2 from the atmosphere and trap it in a frozen, solid form.	
Text-based Question	**Inference Question**
What is the greenhouse effect? A. the absorption of CO2 by growing plants **B. the trapping of radiation** C. the increase in heat of the earth due to sunspots D. the increase in burning of fossil fuels	What can cause less solar radiation to reach earth? A. when the Earth's orbit is closer to the Sun B. sunspots C. **the formation of more mountain ranges** D. the seasons

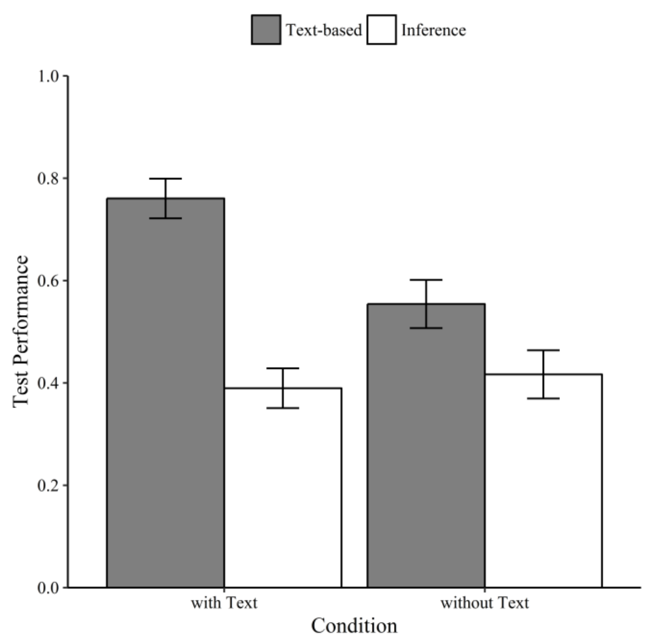

Figure 1. Test performance of readers compared across conditions by question type in Experiment 1 (Error bars represent 95% confidence intervals)

Discussion

The goal of this study was to examine the role that memory for the text plays on test performance with grade-level appropriate texts. The results clearly showed that only performance on text-based questions improved when readers had the text available during testing, conceptually replicating the results seen in Ferrer et al. (2017) and in contrast to the results of Ozuru et al. (2007). This suggests that performance on text-based questions is based on the ability of the reader to maintain information from the text in memory and recall it during testing. However, inference questions were more difficult, and text availability failed to improve performance on inference questions with grade-level appropriate texts.

Experiment 2

A second possible reason for poor learning outcomes when attempting to learn from expository science texts is that students may fail to engage in appropriate reasoning processes. What is critical for the construction of a coherent situation model is that readers go through an active process of generating connections among ideas in the text and between ideas in the text and their prior knowledge. This typically requires a series of causal inferences to integrate pieces of information into an accurate mental model of the phenomena (Graesser, Leon, & Otero, 2002; Kintsch, 1994; Wiley, Griffin, & Thiede., 2005).

There is a substantial body of research that has identified that prompting students to engage in elaboration or explanation while reading is an effective instructional manipulation that can lead to robust improvements in subject-matter learning (Chi, 2000; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; McNamara, 2004; Wiley & Voss, 1999). Explanation activities generally require generative responses and promote constructive processing by prompting a student to ask themselves "how and "why" questions in order to infer the deeper meaning of a passage (Hinze, Wiley, & Pellegrino, 2013). The c*onstructive retrieval hypothesis* (Hinze et al., 2013) proposes that engaging in reasoning processes such as these may be necessary to improve a reader's performance on test questions that tap causal inferences and understanding of the systems or processes introduced within the text.

To test whether *constructive* retrieval processes improve understanding to a greater extent than simply retrieving information from memory, Hinze et al. (2013) had undergraduates read a series of five short science texts written at a middle-school level with three different types of learning activities. They manipulated the level of constructive processing that the reader was prompted to engage in after reading a text through either rereading, a free recall activity, or an explanation activity. Results showed that those who engaged in explanation activities outperformed both the rereading and the free recall groups on both text-based and inference questions. Additionally, it was proposed that the quality of the reasoning processes that participants engaged in during the activities would also predict learning from text. After coding all written responses, it was found that the quality of explanation was predictive of both text-based and inference question performance.

Based on this prior work, Experiment 2 manipulated whether participants engaged in a constructive activity during or after the reading process. Some participants were encouraged to engage in constructive processing by writing short explanations after reading each text. Other participants were encouraged to engage in constructive processing during reading by engaging in think-aloud protocols with explicit prompts to produce explanations embedded within them. These explanation prompts, presented at five strategic points within each text, required readers to engage in reasoning during the reading process. In addition, these students also produced short written explanations after reading each text. The goal of each of these activities was to help readers to construct more coherent situation models of the texts, which should improve performance on inference questions. Further, it was predicted that the combination of prompting students to engage in appropriate reasoning during reading, and constructive retrieval after reading, would provide the most support for learners. Finally, based on the relation between explanation quality and performance found in Hinze et al. (2013), it was predicted that those engaging in high-quality reasoning would show the best performance.

Method

Participants Participants (29 females; M_{age} = 18.8, SD_{age} = 1.0) were 48 undergraduates who received course credit for their participation in the experiment through the introductory psychology subject pool. Participants were randomly assigned to one of two between-participants conditions: writing an explanation after reading each text (explanation) or engaging in think-aloud protocols during reading and writing an explanation after reading (think-aloud). Each participant read and was tested on two of the six texts. Texts were fully counterbalanced so that each of the 6 texts were assigned to 8 participants in each of the 2 conditions, resulting in 96 observations.

Materials Texts and test questions used were identical to those in Experiment 1.

Procedure Participants were randomly assigned to read two of the six topics. Prior to reading each text, participants were instructed to read for understanding. They were told, *"Your goal while reading this text is to develop an understanding of... (how food allergies develop). You will be asked to answer this question after you have finished reading the text, so pay close attention to elements of the text that help you answer this question."* When participants finished reading, they wrote an explanation in response to the question, "*How* *did this information help you to understand… (how food allergies develop)?"* without the text available. After reading and generating an explanation for the first topic, they repeated this procedure for the second topic.

In the think-aloud condition, after participants read through the first text at their own pace, a researcher explained that they would be asked to reread the text in segments; they would be stopped at five points and asked to "*explain what they were thinking and how the current section of the text helped them to understand… (how food allergies develop)."* During the think-aloud protocols, participants received no feedback from the experimenter. Following the think-aloud protocols, they wrote an explanation of the text, and then repeated the procedure for the second text.

After writing the second explanation, participants in both conditions completed the tests without access to the texts.

Explanation Coding Two raters categorized the short written explanations generated by students as being either low or high-quality. As seen in Table 2, low quality explanations were considered to be incoherent, nonsensical, or contained an inaccurate causal assertion. They commonly contained only superficial surface features described in the text. High quality explanations contained an accurate representation of the causal concepts presented in the text. These explanations clearly identified a directional or cause-effect relationship between ideas. Interrater agreement resulted in Cohen's kappa of .81.

Table 2.
Examples of Explanation Quality

Low-Quality Explanation	High-Quality Explanation
The ice ages occur because of the temperature changes. The water levels are low and the temperature in the air is cold. The earth is cold for a period of time and then warm for a shorter period of time.	Ice ages stop when there is a warming period. The warming period happens because of CO2 gases being produced and when there are more CO2 gases, radiation is trapped in the atmosphere, making the earth hotter.

Results

A 2 (Condition: Explanation, Think-Aloud) x 2 (Explanation Quality: High Low) x 2 (Question Type: Text-based, Inference) repeated measures ANOVA indicated no main effect of condition, $F < 1$, but a main effect of question type, $F(1,92) = 13.43$, $p < .001$, $\eta^2 = .13$. As shown in Figure 2, inference questions were generally more difficult that text-based questions.

There was also a main effect for explanation quality, as students who wrote higher quality explanations performed better on test questions, $F(1,92) = 5.68$, $p < .02$, $\eta^2 = .06$. These significant main effects were subsumed by three significant two-way interactions (Condition x Explanation Quality: $F(1,92) = 4.35$, $p < .04$, $\eta^2 = .05$; Question Type x Condition: $F(1,92) = 5.18$, $p < .03$, $\eta^2 = .05$; Question Type x Explanation Quality: $F(1,92) = 4.32$, $p < .04$, $\eta^2 = .05$). The three-way interaction did not reach significance, $F(1,92) = 1.19$, $p < .28$, $\eta^2 = .01$.

Follow-up tests to explore the significant two-way interactions were performed for each question type separately. Starting first with performance on inference questions, as shown in Figure 3 a follow-up 2 (Condition) x 2 (Explanation Quality) ANOVA resulted in a significant interaction, $F(1,92) = 6.30$, $p = .01$, $\eta^2 = .06$. There was also a significant main effect of explanation quality, $F(1,92) = 7.49$, $p = .007$, $\eta^2 = .08$, which was subsumed by the interaction. No main effect of condition was found, $F <1$. Planned comparisons showed that for those writing low-quality explanations, think-aloud prompts during reading significantly affected performance on inference questions over solely explaining the text after reading, $t(92) = 2.38$, $p = .02$, $d = .89$. No differences across conditions were observed for those having written high-quality explanations, $t < 1$.

As shown in Figure 4, the same 2 x 2 ANOVA was conducted for performance on text-based questions. This resulted in no main effect due to explanation quality, $F < 1$, and no interaction, $F = 1.14$. The main effect for condition was not significant, $F(1,92) = 2.47$, $p = .12$, $\eta^2 = .03$, but trended toward better performance on text-based questions in the written explanation only (without think-aloud) condition.

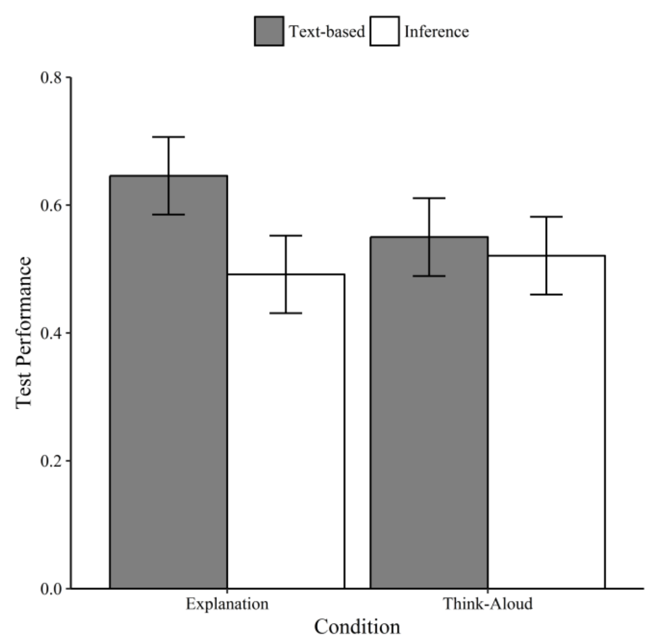

Figure 2. Test performance of readers compared across conditions by question type in Experiment 2 (Error bars represent 95% confidence intervals)

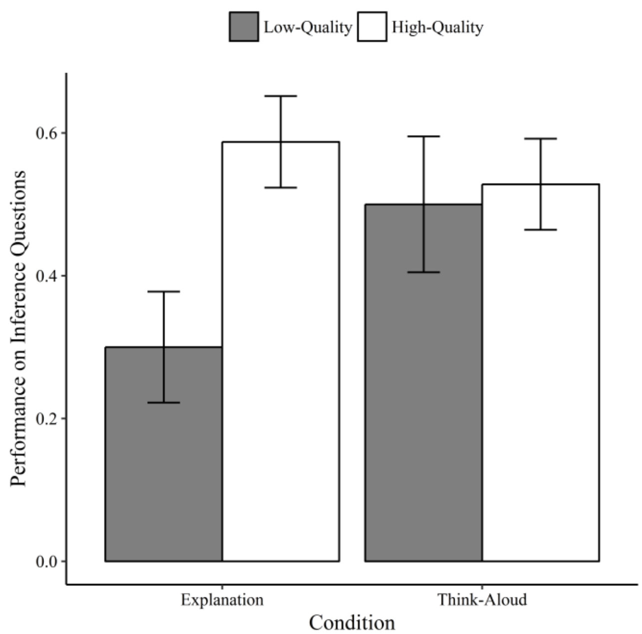

Figure 3. Performance on inference questions across conditions and explanation quality level in Experiment 2 (Error bars represent 95% confidence intervals)

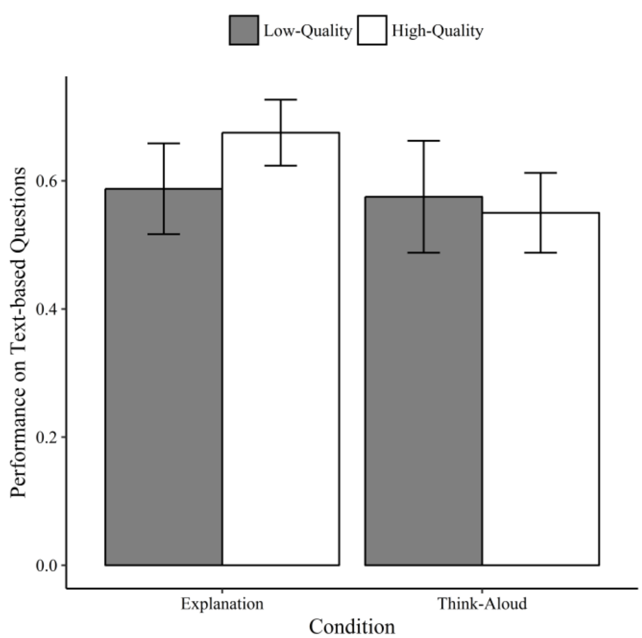

Figure 4. Performance on text-based questions across conditions and explanation quality level in Experiment 2 (Error bars represent 95% confidence intervals)

Discussion

The main goal for this study was to test whether increasing the level of reasoning that a reader was prompted to engage in during and after reading would improve test performance. Advantages in performance on inference questions were seen for participants who were prompted to engage in constructive processes both during and after reading. However, no benefits of the added reasoning activities were seen for text-based questions indicating that text-based questions do not rely heavily on the reader engaging in appropriate reasoning processes. Further, the added think-aloud prompts only benefitted inference test performance for participants who were writing low-quality explanations. Performance did not differ for those who wrote high-quality explanations indicating that those readers were likely already engaging in the appropriate reasoning processes without the need for additional scaffolds embedded within the think-aloud.

General Discussion

The main purpose of these studies was to explore two possible reasons why students struggle with comprehension from expository science texts. The first possibility was that students suffer from poor memory for the texts. To test this hypothesis, the availability of the text during testing was manipulated. Prior research showed that simply giving readers access to text during testing improved performance on both text-based and inference questions when participants were reading below their grade level (Ozuru et al., 2007). However, when participants were given a grade-level appropriate text, only performance on text-based questions increased with access to the text during testing (Ferrer et al., 2017). Consistent with predictions based in Ferrer et al. (2017), a significant difference in performance was seen in Experiment 1 for text-based questions only.

This dissociation in performance on text-based and inference questions in Experiment 1 also provides validation for how these two types of questions were originally designed (Wiley et al., 2005). The text-based questions were designed so that answers could be found directly in text. The inference questions were created with the explicit intent of measuring a reader's ability to integrate information and to construct a coherent situation model, mental model, or causal model of the system or process being described by the text. That is, answers to inference questions were not readily accessible in the text using the same method of verbatim search that could be used for text-based questions. Experiment 1 showed that the inference questions were more difficult for students to answer, and also that performance on inference questions did not seem to depend on memory for the text.

This leads to the second possibility that was considered in Experiment 2: that performance on inference questions depends on the quality of reasoning that a reader engages in during the reading process. Prior research has shown that the addition of constructive activities can improve learning outcomes from expository science texts. Experiment 2 showed a benefit of prompting reasoning both during and after reading on inference question performance, and it was particularly the participants who wrote low-quality explanations that needed this support.

Although prior work has improved performance on both memory and inference questions with the same manipulation, here dissociations were found such that text availability only altered performance on text-based questions, while the think-aloud manipulation only improved performance on inference questions. One salient difference between prior research that failed to find dissociations between question types and the current studies is in the complexity of the texts. Inference questions may be especially challenging when students are learning from difficult texts, and it may be under these contexts that memory for text and understanding from text may be most likely to diverge.

Conclusion

These findings showed that having the ability to reference the text during testing is sufficient for improving performance on text-based questions; while improvements on inference questions may require conditions that support readers in engaging in appropriate reasoning processes during study. The overarching goal of the current work is to understand how students can be assisted in developing a deeper understanding of ideas presented within a text.

This work suggests that struggling readers may benefit from additional scaffolds to help them to generate accurate and appropriate inferences when the goal for reading is to build a coherent causal model of systems, processes or phenomena from complex expository texts. Theoretically, the construction of these models should allow students to transfer this knowledge to new contexts. One important direction for future research is to test this assumption with delayed tests. Additionally, it would be useful to replicate the current experiments within actual classroom contexts to see the effects of the manipulations in a higher-stakes environment.

A broader point is that the results found in these experiments help to reinforce the important differences that need to be acknowledged between memory for a text and developing understanding from a text (Kintsch, 1994). There is a wide variety in the types of items used in standardized comprehension tests, by teachers in classroom contexts, as well as by researchers who conduct studies of learning from text (Wiley & Guerrero, in press). Some may include only text-based or verbatim memory questions. Some may emphasize inference questions. Many may include a mix of different types of questions. Given the dissociations seen here between performance on text-based and inference questions, this suggests that one needs to carefully consider the extent to which a test is assessing memory versus understanding of a text. Which conditions or activities are best for student learning is likely to depend on the goal of instruction.

Acknowledgments

The authors were supported by grants from the National Science Foundation (DUE 1535299) and the Institute of Education Sciences (R305A160008) during the preparation of this manuscript. All opinions expressed herein are those of the authors and do not necessarily reflect those of the funding agencies.

References

Chi, M.T.H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. *Advances in Instructional Psychology*, *5*, 161-238.

Dunlosky, J., Rawson, K.A., Marsh, E.J., Nathan, M.J., & Willingham, D.T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*, 4-58.

Ferrer, A., Vidal-Abarca, E., Serrano, M.Á., & Gilabert, R. (2017). Impact of text availability and question format on reading comprehension processes. *Contemporary Educational Psychology, 51, 404-415.*Graesser, A.C., & Bertus, E.L. (1998). The construction of causal inferences while reading expository texts on science and technology. *Scientific Studies of Reading, 2*, 247-269.

Graesser, A.C., Leon, J.A., & Otero, J. (2002). Introduction to the psychology of science text comprehension. In J. Otero, J.A. Leon, & A.C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 1-15). Mahwah, NJ: Lawrence Erlbaum.

Hinze, S.R., Wiley, J., & Pellegrino, J.W. (2013). The importance of constructive comprehension processes in learning from tests. *Journal of Memory and Language*, *69*, 151-164.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*, 294-303.

Mayer, R. (1989). Models for understanding. *Review of Educational Research, 59*, 43-64.

McNamara, D.S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38,* 1-30.

Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D.S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, *25*, 399-438.

Thiede, K.W., Wiley, J., & Griffin, T.D. (2011). Test expectancy affects metacomprehension accuracy. *British Journal of Educational Psychology, 81,* 264-273.

Wiley, J., Griffin, T.D., & Thiede, K.W. (2005). Putting the comprehension in metacomprehension. *Journal of General Psychology, 132*, 408-428.

Wiley, J., & Guerrero, T.A. (in press). Prose comprehension beyond the page. To appear in K. Millis, J. Magliano, D. Long, & K. Wiemer (Eds). *Deep learning: Multi-disciplinary approaches.* Routledge/Taylor and Francis.

Wiley, J., & Voss, J.F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91,* 301-311.