Paper Title Generating Explanations Is More Helpful Than Practice Testing Alone for Improving Comprehension and Metacomprehension

Author(s) Tricia Ann Guerrero, University of Illinois at Chicago; Thomas D. Griffin, University of Illinois at Chicago; Jennifer Wiley, University of Illinois at Chicago

Session Title Motivation and Self-Regulation in Technology-Based Environments

Session Type Poster Presentation

Presentation Date 4/19/2020

Presentation Location Online

Descriptors Learning Environments, Metacognition, Teaching and Learning

Methodology Quantitative

Unit Division C - Learning and Instruction

DOI https://doi.org/10.3102/1578055

This paper should be cited as:
Guerrero, T. A., Griffin, T. D., & Wiley, J. (2020, April). *Generating explanations is more helpful than practice testing alone for improving comprehension and metacomprehension.* Paper presented at the 2020 Annual Meeting of the American Educational Research Association. Retrieved [Retrieval Date], from the AERA Online Paper Repository. https://doi.org/10.3102/1578055

AERA 2020 Online Paper Repository

Generating explanations is more helpful than practice testing alone for improving comprehension and metacomprehension

Tricia A. Guerrero, Thomas D. Griffin, Jennifer Wiley
University of Illinois at Chicago

Objectives or purposes

The main goal for this study was to test whether online homework systems including practice testing activities alone can improve comprehension and skills in comprehension monitoring, or whether engaging in explanation activities is critical for better outcomes within the context of a gateway science course that is generally taken by students as they first enter college.

Perspectives and theoretical framework

Although being able to self-regulate comprehension behaviors is critical for students to learn effectively from expository science text, students are generally poor at monitoring their own understanding (Maki, 1998; Thiede, Griffin, Wiley, & Redford, 2009; Winne & Hadwin, 1998). They are overconfident in their ability to understand phenomena after first reading an expository science text, and poor at distinguishing which topics they have understood most from topics they understood least. However, activities such as generating summaries or keywords after reading can help students to engage in more accurate comprehension monitoring (Thiede, Dunlosky, Griffin, & Wiley, 2005). Research has also found that prompting students to generate concept maps, sketches, or explanations during reading can improve metacomprehension accuracy (Fukaya, 2013; Griffin, Wiley, & Thiede, 2008, 2019; Jaeger & Wiley, 2014; Thiede et al., 2010; van Loon, de Bruin, van Gog, van Merriënboer, & Dunlosky, 2014; Wiley, 2019). While most of these effects have been found within a laboratory setting, an open question is whether these benefits might also be seen in a gateway science course context.

Alternatively, work on "testing effects" has suggested that repeated attempts at taking practice tests on the same topics (and thus engaging in retrieval practice) can lead to robust learning (Adesope, Trevisan, & Sundararajan, 2017; Rawson & Dunlosky, 2012; Roediger & Karpicke, 2006). Although most work on "testing effects" has focused on improving skill acquisition, or memory for discrete units of information such as facts or definitions, some researchers have shown that benefits can be seen from repeated tests on the same content even when students are tasked with understanding concepts from expository science texts (Jensen, McDaniel, Woodard & Kummer, 2014; Johnson & Mayer, 2009; Karpicke & Blunt, 2011). In contrast, when comprehension monitoring outcomes are considered, Griffin, Wiley, and Thiede (2019) found that taking practice tests on one set of topics did not improve students' ability to predict their performance on a new set of topics, unless the practice tests were accompanied by a general description about the type of test items to expect (memory versus inference). It remains to be seen whether taking practice tests might improve metacomprehension skills, particularly within an authentic course context.

Thus, the main focus of the present study assessed differences in metacomprehension outcomes as a function of two common study activities – practice testing and generating explanations – which were

integrated into an online homework system within the context of a gateway science course that is generally taken by students in their first year of college.

Methods, techniques or modes of inquiry

Students (*N* = 308) enrolled in Introduction to Psychology at a large, diverse university completed online homework activities over the first several weeks of the course in either an explanation-generation or practice-testing condition. The homework activities used textbook excerpts on 6 key topics (Fundamental Attribution Error, Cognitive Dissonance, Conformity/Obedience to Authority, Self-Control, Expectancy Effects, Representative Sampling). The textbook excerpts were approximately 750 words each (Flesch-Kincaid grade level 12). The test questions were multiple choice (consistent with the actual exams for the course) and were developed in consultation with the instructor to test understanding of each of the theories or phenomena. Each question required the reader to make inferences, understand implicit connections among ideas, and be able to apply the concepts from the text to new situations. They were not verbatim memory questions.

During the first homework activity, students read the textbook excerpts at their own pace. Then they provided predictive judgments of how well they thought they would do on a 5-item quiz testing their understanding of the content of each excerpt (Judgments of Comprehension, JOCs, 0-5 scale, converted to proportions out of 5 to match test scores), and selected which text they would be most likely to restudy. Students then took 5-item practice tests on each topic (scored as proportions correct). They were informed of the correct responses after completing all of the practice tests. During the next two homework activities, all students had a chance to re-read the excerpts and to attempt the same practice test items as they had seen in the first homework again. However, in addition to attempting the practice items over again, students in the explanation-generation condition were prompted to engage in generating their own explanations of how the evidence provided in the text (examples and results of empirical studies) supported the theories that were being described. They were encouraged to notice connections among the parts of each text in order to be able to answer inference questions. In a fourth homework activity, all students took a new set of tests with 5 new items (with new stems and response alternatives) for the same topics. By comparing scores on the new tests (from Homework 4) to scores on the initial tests (from Homework 1), this provided a measure of comprehension gains due to engaging with content during the two different homework activities.

After the instructional activities, all students also completed a transfer unit as a final homework activity. To assess whether the instructional activities led to transferable comprehension monitoring skills, the metacomprehension measures needed to be based on predictions and performance for a new set of topics. Thus, students were given excerpts for a new set of 6 topics to study, made predictive judgments about how well they thought they would do, selected a text for restudy, and then took 5-item tests on their understanding of each topic. The 6 new topics were Classical Conditioning, Operant Conditioning, Observational Learning, Aphasia, Confirmation Bias, and Intelligence Testing. The transfer unit provided the opportunity to look at changes in metacognitive skills compared to the baseline skills assessed in Homework 1.

Data sources, evidence, objects or materials

Comprehension gains were tested with a mixed ANOVA comparing the two conditions in terms of gains in average scores from the Homework 1 tests to the Homework 4 tests on the same topics that students studied repeatedly as part of the homework activities.

The remaining measures were measures of metacognitive skills, derived by comparing the transfer unit

to the Homework 1 baseline assessment. The changes in three measures were tested using mixed ANOVAs comparing the two conditions. Confidence bias was computed as the difference between average JOCs and actual test performance (positive values mean overconfidence, and more extreme values represent more error in prediction). Relative metacomprehension accuracy was computed as the intra-individual correlation between each student's JOCs and comprehension test performances across the set of topics. For this measure, higher values mean a student was better able to monitor their comprehension across topics and had a more accurate sense of which topics they understood best or worst. In addition, a measure of regulation was also computed by categorizing re-study choices as good or poor with respect to whether students actually performed worst on the topic that they identified as needing to restudy the most.

Results

Comprehension gains. Regardless of condition (either practice-testing or explanation-generation activities), all students showed significant gains from the initial Homework 1 tests to the later Homework 4 tests on the same topics. As shown in Figure 1, the explanation condition tended to gain more, however the difference in gains between conditions was not significant.

Changes in metacognitive skills. In general, all students tended to be overconfident on the baseline assessment given during Homework 1. As shown in Figure 1, overconfidence increased from baseline to the transfer assessment. However, overconfidence increased more for students in the practice-testing condition.

Average relative accuracy was near zero on the baseline assessment. However, as shown in Figure 1, relative accuracy increased on the transfer set for students who completed the explanation activities, while it decreased for students who completed only practice tests.

Finally, as shown in Figure 1, a similar proportion of students in the practice-testing condition chose to restudy the text they understood the least during Homework 1 and on the transfer set. However, for students in the explanation generation condition, a significantly greater proportion of students selected the least understood topic for restudy on the transfer set.

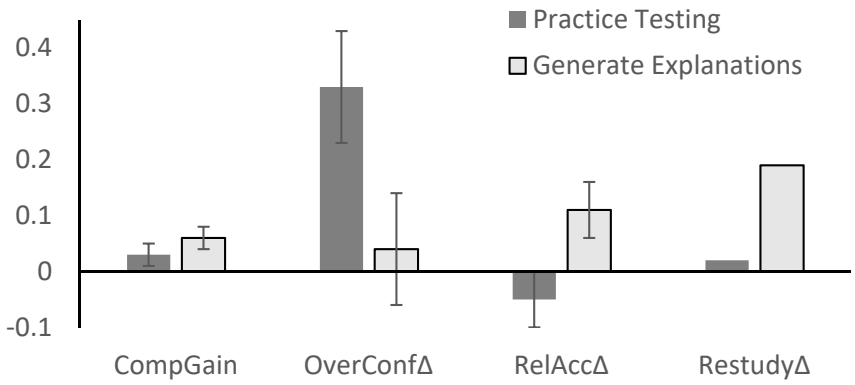

Figure 1. Effects of activity conditions on average gains in comprehension test scores (CompGain), average changes in overconfidence (OverConfΔ), average changes in relative accuracy (RelAccΔ), and increases in frequencies of good restudy choices (RestudyΔ) from before to after engaging in different homework activities. Error bars represent standard errors.

Scholarly and scientific significance

In summary, online homework activities that involved practice testing alone facilitated comprehension of the studied topics, but they failed to improve skills in comprehension monitoring and exacerbated overconfidence. Meanwhile, online homework activities that involved explanation generation not only facilitated comprehension, they also helped students to moderate their overconfidence. Further, it was only when students were prompted to engage in explanation activities as part of their online homework that relative metacomprehension accuracy increased on the transfer test. In addition, only the explanation activities led to improvements in restudy choices. This suggests that these explanation-based instructional activities, integrated as part of an online homework system, are giving students a better sense of what they need to study further as they engage in self-regulating their learning in this gateway science course.

These results are consistent with prior work showing metacognitive benefits from attempting to generate explanations while studying expository science texts (Fukaya, 2013; Griffin, et al. 2008, 2019; Jaeger & Wiley, 2014), and help to extend them into a real course context. They are also consistent with work showing that practice tests alone are insufficient to help students develop appropriate expectations that allow them to more accurately monitor their relative levels of comprehension on new material (Griffin et al., 2019). With respect to this measure, it is important to note that the baseline levels of relative accuracy were quite low compared to average levels of studies found in laboratory contexts (Dunlosky & Lipko, 2007). The present results are consistent with the recent observation that relative accuracy tends to be particularly low in real course contexts, likely due to the high conceptual overlap between the different texts which makes it more difficult to make relative judgments of comprehension (Griffin et al., 2019). This also underscores the importance of developing instructional approaches, such as this online homework system, to help students to improve their metacomprehension skills under the more stringent constraints within a real course context.

In contrast, the comprehension outcomes generally failed to show the advantages that are often found from explanation activities (Chi, 2000; Guerrero & Wiley, 2018; Hinze, Wiley, & Pellegrino, 2013; McNamara, 2004). However, Hinze et al. (2013) found that explanation activities were more likely to improve performance on new comprehension items only if students engaged in constructive retrieval that went beyond simple recall of concepts that had been explicitly stated in the text. Ongoing work is exploring how to further enhance the benefits students might gain from the explanation activities by attempting to provide students with feedback about the quality of their explanations as part of the online homework system (Guerrero & Wiley, 2019). It is hoped that providing this additional support might also help to enhance the effects of the instructional activities on improving students' monitoring skills.

References

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*, 659-701. https://doi.org/10.3102/0034654316689306

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161-238). Erlbaum.

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*, 228-232. https://doi.org/10.1111/j.1467-8721.2007.00509.x

Fukaya, T. (2013). Explanation generation, not explanation expectancy, improves metacomprehension accuracy. *Metacognition and Learning*, *8*, 1-18. https://doi.org/10.1007/s11409-012-9093-0

Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition, 36,* 93-103. https://doi.org/10.3758/MC.36.1.93

Griffin, T. D., Wiley, J., & Thiede, K. W. (2019). The effects of comprehension-test expectancies on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45,* 1066-1092. https://doi.org/10.1037/xlm0000634

Guerrero, T. A., & Wiley, J. (2018). Effects of text availability and reasoning processes on test performance. In T. Rogers, M. Rau, X. Zhu, & C. Kalish (Eds.), *CogSci 2018 Proceedings* (pp. 1748-1753). https://cogsci.mindmodeling.org/2018/papers/0336/0336.pdf

Guerrero, T. A. & Wiley, J. (2019). Using "idealized peers" for automated evaluation of student understanding in an introductory psychology course. In S. Isotani, E. Millan, & A. Ogan (Eds.), *Artificial Intelligence in Education – 20th International Conference, AIED2019* (pp. 133-143). Springer Nature. https://doi.org/10.1007/978-3-030-23204-7_12

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. https://doi.org/10.1016/j.jml.2013.03.002

Jaeger, A. J., & Wiley, J. (2014). Do illustrations help or harm metacomprehension accuracy? *Learning & Instruction, 34*, 58-73. https://doi.org/10.1016/j.learninstruc.2014.08.002

Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test… or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, *26*, 307-329. https://doi.org/10.1007/s10648-013-9248-9

Johnson, C. I. & Mayer, R. E. *(*2009). A testing effect with multimedia learning. *Journal of Educational Psychology, 101*, 621-629. https://doi.org/10.1037/a0015183

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*, 772-775. https://doi.org/10.1126/science.1199327

Maki, R. H. (1998). Metacomprehension of text: Influence of absolute confidence level on bias and accuracy. In D. L. Medin (Ed.), *The psychology of learning and motivation* (pp. 223-248). Academic Press.

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38*, 1-30. https://doi.org/10.1207/s15326950dp3801_1

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249-255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rawson, K. A., & Dunlosky, J. (2012). When is practice testing most effective for improving the durability and efficiency of student learning? *Educational Psychology Review*, *24*, 419-435. https://doi.org/10.1007/s10648-012-9203-1

Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005). Understanding the delayed-keyword effect on metacomprehension accuracy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1267-1280. https://doi.org/10.1037/0278-7393.31.6.1267

Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds), *Handbook of metacognition in education* (pp. 85-106). Routledge.

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes, 47*, 331-362. https://doi.org/10.1080/01638530902959927

van Loon, M. H., de Bruin, A. B., van Gog, T., van Merriënboer, J. J., & Dunlosky, J. (2014). Can students evaluate their understanding of cause-and-effect relations? The effects of diagram completion on monitoring accuracy. *Acta Psychologica, 151*, 143-154. https://doi.org/10.1016/j.actpsy.2014.06.007

Wiley, J. (2019). Picture this! Effects of photographs, diagrams, animations, and sketching on learning and beliefs about learning from a geoscience text. *Applied Cognitive Psychology, 33*, 9-19. https://doi.org/10.1002/acp.3495

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated engagement in learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Lawrence Erlbaum.

Acknowledgements

This research was supported by a grant from the Institute of Education Sciences (R305A160008).