



## Test expectancy affects metacomprehension accuracy

Keith W. Thiede<sup>1\*</sup>, Jennifer Wiley<sup>2</sup> and Thomas D. Griffin<sup>2</sup>

<sup>1</sup>Boise State University, Boise, Idaho, USA

<sup>2</sup>University of Illinois at Chicago, Illinois, USA

**Background.** Theory suggests that the accuracy of metacognitive monitoring is affected by the cues used to judge learning. Researchers have improved monitoring accuracy by directing attention to more appropriate cues; however, this is the first study to more directly point students to more appropriate cues using instructions regarding tests and practice tests.

**Aims.** The purpose of the present study was to examine whether the accuracy of metacognitive monitoring was affected by the nature of the test expected.

**Sample and method.** Students ( $N = 59$ ) were randomly assigned to one of two test expectancy groups (memory vs. inference). Then after reading texts, judging learning, completed both memory and inference tests.

**Results.** Test performance and monitoring accuracy were superior when students received the kind of test they had been led to expect rather than the unexpected test.

**Conclusion.** Tests influence students' perceptions of what constitutes learning. Our findings suggest that this could affect how students prepare for tests and how they monitor their own learning.

Accurate monitoring of learning (i.e., differentiating material that is learned from that which is not) is critical to effective regulation of study and learning (Thiede, Anderson, & Therriault, 2003). However, monitoring accuracy is often quite poor for complex learning tasks, such as learning from texts (Dunlosky & Lipko, 2007; Maki, 1998; Thiede, Griffin, Wiley, & Redford, 2009). Further, past research has shown that monitoring accuracy is particularly poor when students are asked to judge their comprehension from expository texts, and when comprehension is assessed with tests that require the generation of inferences about what was read, rather than memory for what was read (Weaver & Bryant, 1995; Wiley, Griffin, & Thiede, 2005).

Nearly, all teachers can recall a conversation with a student lamenting poor performance on a test in which the student states, 'I can't believe I did so poorly on the test. I really knew this stuff'. Why are students so far off the mark? The prevailing

\*Correspondence should be addressed to Dr Keith W. Thiede, Boise State University, MS 1745, 1910 University Drive, Boise, ID 83725-1745, USA (e-mail: keiththiede@boisestate.edu).

theory used to explain poor monitoring accuracy is that readers are using the wrong cues to monitor their learning (Koriat, 1997). The observation that students may have particular difficulty judging their comprehension is consistent with findings that students often struggle to identify important information in to-be-learned material (e.g., Brown & Smiley, 1977). Other studies have found that readers tend to rely on their surface memory for a text as a default when making comprehension judgments (Thiede, Griffin, Wiley, & Anderson, 2010). If this is the case, then when tests require students to identify important ideas or make inferences, readers' predictive judgments of how they will perform could be quite inaccurate. The present study explores the assumption that poor metacomprehension accuracy may be in part due to a lack of understanding of what comprehension means, or in practical terms, what students think they will be tested on following reading. Following this reasoning, one strategy that may be useful for improving monitoring accuracy might be to give students information about the kind of questions that will be asked, and to see whether they can then use that information to adjust their monitoring of their own learning accordingly. This study explored whether we can improve monitoring accuracy by giving students practice tests on a different set of texts.

There are two lines of prior research that relate to the present question which we will characterize as test 'congruency' studies, and test expectancy studies. The first area of work consists of a growing number of studies that suggest that congruency between encoding and testing conditions, or judging and testing conditions, is critical for accurate monitoring. For example, several studies have shown that when readers make comprehension judgments at a delay, and after attempting to generate keywords or summaries, they can produce more accurate judgments of their later test performance (Thiede & Anderson, 2003; Thiede *et al.*, 2003; Thiede, Dunlosky, Griffin, & Wiley, 2005). These studies have incorporated normal reading of texts, but have attempted to improve monitoring accuracy through manipulations to the judgment process - attempting to create congruency between judgments and tests. Similarly, Thomas and McDaniel (2007) showed that readers can have good monitoring accuracy when later tests are well matched with initial encoding activities. In this study, the initial encoding of texts was manipulated by presenting the to-be-understood texts either with missing letters or in a scrambled order. The scrambled presentation required sentence sorting at encoding, and this was designed to focus readers on the connections among ideas contained in texts. The missing letter presentation required letter insertion during reading of the texts, and this was designed to focus readers on the surface details of the texts (McDaniel, Hines, Waddill, & Einstein, 1994). The findings from this study showed two congruency effects. First, readers did better on test performance when encoding matched testing. In addition, the study showed a congruency effect for monitoring accuracy. That is, the sentence-sorting task improved monitoring accuracy when the tests assessed conceptual understanding - but not when tests assessed memory of details. And, the letter insertion task improved monitoring accuracy when tests assessed memory of details - but not when the tests assessed conceptual understanding. Although these findings are interesting for theoretical reasons, a concern with applying them to actual classroom situations is that these encoding conditions are highly artificial, and a very important practical question for teachers is how we can improve monitoring accuracy under more naturalistic reading conditions. The proposition tested in the current research is that instilling a test expectancy in students before reading might lead to better congruency between encoding, judging, and testing; and this should improve monitoring accuracy.

Following in this vein, the second line of related research is test expectancy studies where students are simply told before learning what sort of test they should expect. The typical finding from test expectancy research, often set in the classroom, is that students do better on tests when they receive the kind of test they expected (for a meta-analysis of this literature, see Lundeberg & Fox, 1991). However, the studies included in this literature are generally examining the effects of expecting a particular test format (recognition vs. recall). We know of no studies that have examined how the levels of processing required by test questions might be varied to set up specific expectations. Nevertheless, it still could be predicted that readers should do better when a test matches their expectations, presumably because they can engage in more effective study behaviours. Still, little is known about how instilling any test expectancy might affect monitoring behaviours (Thiede, 1996). More important, no studies have examined the effects of test expectancy on the accuracy of metacognitive monitoring.

### **Overview of experiment**

In this investigation, we used a test expectancy paradigm to evaluate whether instilling a particular test expectancy affects monitoring accuracy. Participants were asked to read an initial set of texts, and were given practice with *either* detail questions or inference questions. These practice test items were designed to produce expectancy for a particular kind of test and to focus subsequent metacognitive judgments on a particular kind of learning (memory of details vs. inference making). After the practice set, participants read a new set of six critical texts, predicted their test performance, and completed *both* tests that matched their expectations and tests that did not match their expectations for each text. The main question was whether the expectancy, set up during the practice texts, would transfer to a new set of texts and affect monitoring accuracy on the new set. Demonstrating that practice test items that vary in the level of processing they require have a differential effect on later monitoring performance would be a novel finding. If instilling a test expectancy can help to define comprehension for participants and direct them to use the correct information when making their predictions of subsequent test performance on a new set of texts, then we should see improvements in monitoring accuracy.

## **Method**

### **Participants**

Fifty-nine students enrolled in a graduate-level education course participated in the study. Participants completed the experimental session in groups of 4–8 students, with each participant situated at a separate table. All participants were treated in a manner consistent with the ethical standards of the American Psychological Association.

### **Materials**

The texts were expository texts that described a complex causal relation. As suggested by Wiley *et al.* (2005), the texts were developed so that important elements of the situation model were not explicit in the surface form of the text (i.e., the causal connections among ideas in the texts were not stated and needed to be generated by the reader). Each text was approximately 1,000 words long, with an average Flesch–Kincaid readability score

of 11.8. For each text, we constructed two kinds of tests, each containing five questions (an excerpt from a text and sample test questions can be found in the Appendix). One kind of test was designed to assess memory of the details explicitly stated in the text. The other required participants to draw inferences about the ideas presented in the text and was designed to tap the situation model of the text.

### **Design**

Participants were randomly assigned to either the memory expectancy group or the inference expectancy group. Expectancy was manipulated by instructions and practice tests. For critical texts, the nature of the test (memory vs. inference) was manipulated within participants (i.e., participants took both kinds of tests for each text), the order of test type (expected vs. unexpected) was counter-balanced across participants. Thus, the experiment used a 2 (expectancy: memory vs. inference)  $\times$  2 (nature of test: memory vs. inference) design.

### **Procedure**

All the participants were instructed that they would be reading a series of texts, judging their comprehension of each text, and then taking a test for each text. Throughout the experiment, participants were given as much time as needed to read texts, make judgments, and complete tests. Prior to reading the practice texts, they were given instructions regarding the nature of the tests and informed that three texts and tests would be provided for practice. Participants in the memory expectancy group were instructed, 'You will be taking tests that assess your ability to remember specific information contained in the texts'. Participants in the inference group were instructed, 'You will be taking tests that assess your ability to make connections between the different parts of a text (i.e., link the parts of the text)'. For the three practice texts, participants read a text and immediately judged their comprehension. The prompt for the metacognitive judgment was as follows: 'You will take a test of the material you just read. How many questions out of 5 do you think you can correctly answer?' After making their judgment, participants completed a five-item multiple choice test. The nature of each practice test matched the test described in the instructions.

After completing the three practice texts, participants were instructed that they would repeat a similar procedure for six additional texts, the only difference being that the test for each text would be given after reading and judging comprehension of all the texts. Thus, participants read and immediately judged their comprehension of each of the six critical texts. The order of the critical texts was randomized for each participant. Each participant then completed both memory and inference tests. The tests were given in two blocks after the last metacognitive judgment. For half of the participants, the first block of six tests matched their test expectancy (i.e., memory expectancy - memory test; inference expectancy - inference test). The second block of six tests did not match their expectancy (i.e., memory expectancy - inference test; inference expectancy - memory test). For the other half, the blocks were in the opposite order. The title of the text being tested was presented at the top of the test during testing. Although the order of texts was randomized for each participant text presentation order was held constant across reading and both blocks of testing for each participant.

## Results

### Metacognitive judgments and test performance

The primary focus on this investigation is monitoring accuracy; however, as monitoring accuracy is the relationship between metacognitive judgments and test performance, we first report data on these variables.

For each participant, we computed the mean metacognitive judgment across the six critical texts. The mean of the means was computed across participants in each group (see Table 1). The mean magnitude of metacognitive judgments did not differ across groups,  $t(30) < 1.00$ ,  $p = .45$ .

**Table 1.** Descriptive statistics on metacognitive judgments and test performance by expectancy groups

Expectancy group	Judgment magnitude	Memory test performance	Inference test performance
Memory	3.12 (.18)	0.69 (.02)	0.57 (.02)
Inference	3.28 (.11)	0.51 (.02)	0.66 (.02)

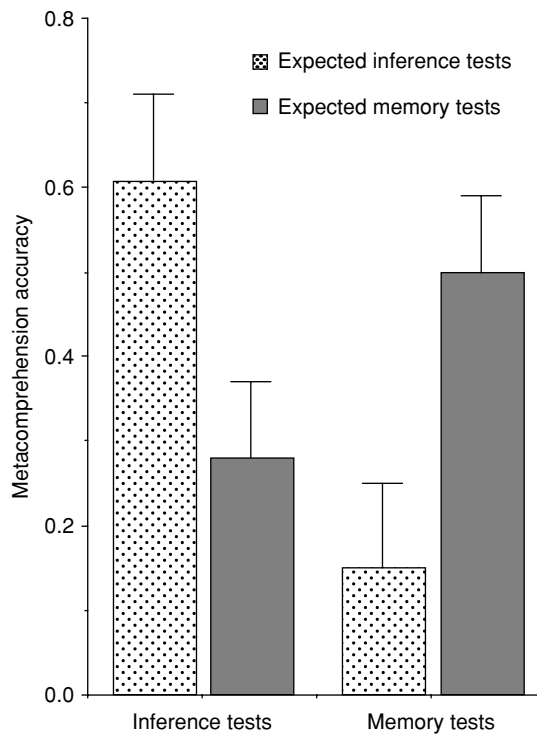
Note. The entries are the mean metacognitive judgment and test performance computed across participants within each condition. The numbers in parentheses are the standard errors of the means.

The mean test performance was computed for each participant. The mean of the means was then computed across participants for each kind of test (memory vs. inference). As shown in Table 1, test performance was superior when there was congruency between the kind of test expected and the kind of test received. There was a significant interaction between test expectancy and the nature of the test,  $F(1, 57) = 44.36$ ,  $MSE = .01$ ,  $p < .001$ ,  $\eta^2 = .44$ . Neither the main effect for test expectancy nor nature of the test was significant, both  $F(1, 57) < 2.0$ ,  $p > .15$ . Follow-up tests of simple effects showed that on inference tests, performance was greater for participants expecting inference tests than for those expecting memory tests,  $t(57) = 2.96$ ,  $p = .004$ ; and on memory tests, performance was greater for participants expecting memory tests than for those expecting inference tests,  $t(57) = 5.40$ ,  $p < .001$ . Moreover, for participants expecting a memory test, performance was greater on memory tests than on inference tests,  $t(28) = 4.23$ ,  $p < .001$ ; and for participants expecting an inference test, performance was greater on inference tests than on memory tests,  $t(29) = 5.20$ ,  $p < .001$ .

### Monitoring accuracy

As recommended by Nelson (1984), monitoring accuracy was operationalized as the intra-individual gamma correlation<sup>1</sup> between a person's metacognitive judgments and test performance across the six critical texts. Two gamma correlations were computed for each participant, one between judgments and performance on memory tests and one between judgments and performance on inference tests. A mean gamma correlation was

<sup>1</sup>Nelson (1984) recommended using a Goodman–Kruskal gamma correlation (Goodman & Kruskal, 1954) for these kinds of data. Gamma is computed by examining the direction of one variable relative to another. If one variable (e.g., metacomprehension judgment) is increasing from one text to another and the other variable (e.g., test performance) is also increasing across this same pair of texts, this is considered a concordance (C). By contrast, if one variable is increasing from one text to another and the other variable is decreasing across this same pair of texts, this is considered a discordance (D). Concordance and discordance is computed across all pairs of items. The total number of each is used to compute the correlation coefficient,  $\gamma = (C - D)/(C + D)$ .



**Figure 1.** Mean monitoring accuracy for each test by expectancy group. The error bars represent the standard error of the mean.

then computed across participants in each expectancy group, for each test (memory and inference). As shown in Figure 1, monitoring accuracy was greater when there was congruency between the expected and actual test than when there was a mismatch between expected and actual test. There was a significant interaction between test expectancy and the nature of the test,  $F(1, 57) = 13.45$ ,  $MSE = .26$ ,  $p < .001$ ,  $\eta^2 = .19$ . Neither the main effect for test expectancy nor nature of the test were significant, both  $F(1, 57) < 1.4$ ,  $p > .24$ . Follow-up tests of simple effects showed that with inference tests, monitoring accuracy was greater for participants expecting inference tests than for those expecting memory tests,  $t(29) = 2.34$ ,  $p = .027$ ; and with memory tests, monitoring accuracy was marginally greater for participants expecting memory tests than for those expecting inference tests,  $t(28) = 2.94$ ,  $p = .007$ . Further, for participants expecting a memory test, monitoring accuracy was greater on memory tests than on inference tests,  $t(57) = 1.73$ ,  $p = .09$ ; and for participants expecting an inference test, monitoring accuracy was greater on inference tests than on memory tests,  $t(57) = 3.17$ ,  $p = .002$ .

## Discussion

Since the early research on transfer-appropriate processing, in the 1970s (e.g., Morris, Bransford, & Franks, 1977), much evidence has been presented to suggest that test performance is enhanced when there is congruency between encoding and testing

conditions (for a review, see Lundeberg & Fox, 1991). Consistent with this literature, we showed a congruency effect for test expectancy. Performance on inference tests was greater for participants expecting inference tests than for participants expecting memory tests; and performance on memory tests was greater for participants expecting memory tests than for participants expecting inference tests. However, unlike other studies that have produced test expectancy effects by creating an expectancy for a particular test format (recall vs. recognition), we obtained the effect by creating an expectancy for tests of differing levels of processing (memory vs. inference). These results are consistent with previous research showing that students adjust their study behaviour in response to the demands of an expected test (McDaniel, Blischak, & Challis, 1994).

We were also able to demonstrate congruency effects on metacognitive monitoring. Monitoring accuracy was greater when there was congruency between the kind of test expected and the kind of test received. Thomas and McDaniel (2007) also showed congruency effects in monitoring; however, a key difference is that we produced the congruency effect with a more naturalistic manipulation. Rather than influence monitoring accuracy by manipulating how materials were encoded by altering the presentation of to-be-learned materials, we influenced monitoring accuracy by exposing participants to a series of practice tests. Participants transferred their knowledge of the tests to a new set of texts, which is a novel finding. Thus, students can apparently adjust their metacognitive monitoring according to the demands of tests.

The above findings are consistent with findings from Kelemen, Winningham, and Weaver (2007) who showed that experience with tests improved monitoring accuracy during an associative learning task (learning Swahili-English translations). However, Maki and Serra (1992) showed only modest improvements in metacomprehension accuracy by presenting practice tests prior to reading – and the modest improvements they found were only evident when practice tests and actual tests matched exactly, not when the practice tests were related to the actual tests (for similar results, see Glenberg, Sanocki, Epstein, & Morris, 1987). One critical difference is that Maki and Serra provided minimal practice with tests; whereas, we provided three practice texts and tests (Kelemen *et al.*, 2007, had participants learn five different lists); therefore, participants had a greater amount of experience with tests in the present studies.

The present findings are also consistent with the transfer-appropriate monitoring (TAM) hypothesis, which states that accuracy of metacognitive monitoring will vary as a function of the match between conditions at the time of the metacognitive judgment and the test. In particular, providing information about the nature of the test may have helped align monitoring to the test, which increased the accuracy of metacognitive monitoring. It is important to note that although there has been some support for TAM in the metacognitive literature (e.g., Begg, Duft, Lalonde, Melinick, & Sanvito, 1989), much more research has produced results inconsistent with TAM (Dunlosky & Nelson, 1997; Dunlosky, Rawson, & Middleton, 2005; Weaver & Kelemen, 2003). However, previous research evaluating TAM has involved less complex learning tasks (e.g., paired-associate learning). As suggested by Weaver and Kelemen (2003), 'It is possible that support for TAM may yet emerge using more complex stimulus materials such as passages of text. Text materials permit a wider range of encoding strategies and processing during judgment and test; this might increase the importance of matching processing at these times' (p. 1064). Our data confirm Weaver and Kelemen's suspicion regarding TAM and monitoring learning from texts. In addition, some other possible factors that may have allowed TAM to be observed in the present study include the combination of providing readers with both instructions and practice test items, the causal nature of the

expository texts, and the design of the test items to reflect either surface memory or inference questions (Griffin, Wiley, & Thiede, 2008; Wiley *et al.*, 2005).

The present findings contribute to both the test expectancy and metacomprehension literatures. This new work is unique in that we showed that instructions and practice tests can be used to produce expectancy for different levels of learning. Our manipulation led some participants to study the details contained in texts and monitor their learning of details; whereas, other participants studied to make inferences across the ideas contained in texts and monitored their learning accordingly. Thus, in contrast to other test expectancy experiments, which have focused on instilling an expectancy for a particular test format (recognition vs. recall), we were able to use a test expectancy paradigm to promote different levels of processing: surface versus deeper learning, and different levels of monitoring.

One educational implication of these findings is that tests influence students' perceptions of what constitutes learning. Years of experience with tests that assess lower levels of learning (recognition of details) may be sending the unintended message to students that learning is more about remembering details than learning deeply and/or learning to apply the concepts presented in school. This could have dramatic effects on the accuracy of metacognitive monitoring.

## Acknowledgements

The research reported here was supported by the Institute of Education Sciences, US Department of Education, through R305B07460 to Keith W. Thiede, Jennifer Wiley, and Thomas D. Griffin. The opinions expressed are those of the authors and do not represent views of the Institute or the US Department of Education. The authors thank Allison Jaeger for her assistance.

## References

- Begg, I., Duft, S., Lalonde, P., Melinick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, *28*, 610-632. doi:10.1016/0749-596X(89)90016-8
- Brown, A. L., & Smiley, S. S. (1977). Rating the importance of structural units of prose passages: A problem of metacognitive development. *Child Development*, *48*, 1-8. doi:10.2307/1128873
- Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science*, *16*, 228-232. doi:10.1111/j.1467-8721.2007.00509.x
- Dunlosky, J., & Nelson, T. O. (1997). Similarity between the cue for judgments of learning (JOL) and the cue for test is not the primary determinant of JOL accuracy. *Journal of Memory and Language*, *36*, 34-49. doi:10.1006/jmla.1996.2476
- Dunlosky, J., Rawson, K. A., & Middleton, E. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language*, *52*, 551-565. doi:10.1016/j.jml.2005.01.011
- Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, *116*, 119-136. doi:10.1037/0096-3445.116.2.119
- Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classification. *Journal of the American Statistical Association*, *49*, 732-764. doi:10.2307/2281536
- 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 and Cognition*, *36*, 93-103. doi:10.3758/MC.36.1.93



- Kelemen, W. L., Winningham, R. G., & Weaver, C. A. (2007). Repeated testing sessions and scholastic aptitude in college students' metacognitive accuracy. *European Journal of Cognitive Psychology, 19*, 689–717. doi:10.1080/09541440701326170
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370. doi:10.1037/0096-3445.126.4.349
- Lundeberg, M. A., & Fox, P. W. (1991). Do laboratory findings on text expectancy generalize to classroom outcomes? *Review of Educational Research, 61*, 94–106.
- Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117–144). Hillsdale, NJ: LEA.
- Maki, R. H., & Serra, M. (1992). Role of practice tests in the accuracy of test predictions on text material. *Journal of Educational Psychology, 84*, 200–210. doi:10.1037/0022-0663.84.2.200
- McDaniel, M. A., Blischak, D. M., & Challis, B. (1994). The effects of test expectancy on processing and memory of prose. *Contemporary Educational Psychology, 19*, 230–248. doi:10.1006/ceps.1994.1019
- McDaniel, M. A., Hines, R. J., Waddill, P. J., & Einstein, G. O. (1994). What makes folk tales unique: Content familiarity, causal structure, scripts, or superstructure? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 169–184. doi:10.1037/0278-7393.20.1.169
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519–533. doi:10.1016/S0022-5371(77)80016-9
- Nelson, T. O. (1984). A comparison of current measures of feeling-of-knowing accuracy. *Psychological Bulletin, 95*, 109–133. doi:10.1037/0033-2909.95.1.109
- Thiede, K. W. (1996). The relative importance of anticipated test format and anticipated test difficulty on performance. *Quarterly Journal of Experimental Psychology A, 49*, 901–918. doi:10.1080/027249896392351
- Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology, 28*, 129–160. doi:10.1016/S0361-476X(02)00011-5
- Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*, 66–73. doi:10.1037/0022-0663.95.1.66
- 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. doi:10.1037/0278-7393.31.6.1267
- Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes, 47*, 331–362. doi:10.1080/01638530902959927
- Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition and self-regulated learning* (pp. 85–106). New York: Routledge.
- Thomas, A. K., & McDaniel, M. A. (2007). The negative cascade of incongruent generative study-test processing in memory and metacomprehension. *Memory and Cognition, 35*, 668–678.
- Weaver, C. A., & Bryant, D. S. (1995). Monitoring of comprehension: The role of text difficulty in metamemory for narrative and expository text. *Memory and Cognition, 23*, 12–22.
- Weaver, C. A., & Kelemen, W. L. (2003). Processing similarity does not improve metamemory: Evidence against transfer-appropriate-monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1058–1065. doi:10.1037/0278-7393.29.6.1058
- Wiley, J., Griffin, T., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *Journal of General Psychology, 132*, 408–428. doi:10.3200/GENP.132.4.408-428

## Appendix

### Example text excerpt and test questions

#### *Excerpt of ice ages passage*

An ice age is a period of time – usually millions or tens of millions of years – when vast glaciers cover as much as a third of the Earth’s land surface. Average global temperatures can drop by as many as 12°C overall. The latest Ice Age began about 2.5 million years ago, and ended approximately 15,000 years ago. Average global temperatures decreased by approximately 8°C. Sea level was lowered substantially due to the amount of water that was frozen in the glaciers. Ice core analysis indicated there were reduced amounts of carbon dioxide in the atmosphere. Giant ice sheets that originated at the North Pole advanced and retreated many times in North America and Europe. The movement of the glaciers coincided with cycles of warm and cold periods in the Earth’s temperature. Throughout history, cycles of changes in global temperatures usually occur every 100,000 years or so. Each cycle consists of a long, generally cold period during which the entire Earth cools, followed by a relatively short warm period during which Earth warms up rapidly.

#### *Sample detail question*

How much of the earth is covered by glaciers during an ice age?

- a. less than 10%
- b. about a third (correct answer)
- c. over half
- d. almost all

#### *Sample inference question*

Higher levels of CO<sub>2</sub> in the atmosphere lead to

- a. higher sea levels (correct answer)
- b. the creation of mountain ranges
- c. the formation of more ice and snow
- d. changes in the earth’s surface