5-1-2010

Poor Metacomprehension Accuracy as a Result of Inappropriate Cue Use

Keith W. Thiede
Boise State University

Thomas Griffin
University of Illinois at Chicago

Jennifer Wiley
University of Illinois at Chicago

Mary C. M. Anderson
College of DuPage

Poor Metacomprehension Accuracy as a Result of Inappropriate Cue Use

Keith W. Thiede

Boise State University

Thomas Griffin & Jennifer Wiley

University of Illinois at Chicago

Mary C. M. Anderson

College of DuPage

running head: Metacomprehension and Cue Use

Send correspondence to: Keith Thiede

Boise State University

Educational Psychology

1910 University Drive

Boise, ID 83725-1745

KeithThiede@boisestate.edu

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Abstract

Two studies attempt to determine the causes of poor metacomprehension accuracy, and then, in turn, to identify interventions that circumvent these difficulties to support effective comprehension monitoring performance. The first study explored the cues that both at-risk and typical college readers use as a basis for their metacomprehension judgments in the context of a delayed summarization paradigm. Improvement was seen in all readers, but at-risk readers did not reach the same level of metacomprehension accuracy as a sample of typical college readers. Further, while few readers reported using comprehension-related cues, more at-risk readers reported using surface-related cues as the basis for their judgments. To support the use of more predictive cues among the at-risk readers, a second study employed a concept map intervention, which was intended to make situation model-level representations more salient. Concept mapping improved both the comprehension and metacomprehension accuracy of at-risk readers. The results suggest that poor metacomprehension accuracy can result from a failure to use appropriate cues for monitoring judgments, and that especially less-able readers need interventions that direct them to predictive cues for comprehension.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Poor Metacomprehension Accuracy as a Result of Inappropriate Cue Use

Learning from text is a standard adjunct to classroom instruction. Students are assigned reading for homework, where they are expected to study and understand textbook chapters or other texts. Models of self-regulated learning (e.g., Dunlosky & Thiede, 1998; Metcalfe, 2002; Nelson & Narens, 1990) suggest that metacognitive monitoring and regulation of study play an important role in such learning. Thiede, Anderson, and Therriault (2003) showed that monitoring accuracy (operationalized as the intra-individual correlation between metacomprehension judgments and test performance computed across texts[1]) influenced decisions about which texts to restudy, which in turn affected learning from text. In particular, they showed that participants who more accurately monitored their comprehension made better decisions about which texts to reread than did participants who less accurately monitored their comprehension. That is, for a group with higher monitoring accuracy, participants chose to restudy primarily the texts that they did not understand. Their mean proportion correct on initial comprehension tests for the texts they selected to reread was .27 versus .78 for the texts they did not select to reread. By contrast, groups with lower monitoring accuracy showed less of a preference. The mean proportion correct on tests for the texts they selected to reread was .43 versus .53 for those they did not select to reread. The more effective regulation of study among the group with higher monitoring accuracy produced higher overall reading comprehension on subsequent tests for that group. Given that these results show that better comprehension monitoring accuracy can lead to better learning from text, it is important to find ways to improve comprehension monitoring accuracy, which has been called *metacomprehension accuracy*.

It is highly problematic, then, that the usual level of metacomprehension accuracy is generally quite dismal, with correlations between predicted test performance and actual

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

performance hovering around .27 (Maki, 1998a, Dunlosky & Lipko, 2007). Prior research has identified a number of constraints that prevent readers from engaging in accurate metacomprehension, but perhaps the most critical one is that readers generally are not basing their judgments on predictive cues for actual comprehension (Rawson, Dunlosky, & Thiede, 2000; Thiede, Wiley, Griffin & Redford, in press; Wiley, Griffin & Thiede, 2005). A great deal of research has been dedicated to identifying the cues that readers use to judge comprehension. This research has suggested that readers use such cues as domain familiarity or interest in the topic (Glenberg & Epstein, 1987; Glenberg, Sanocki, Epstein, & Morris, 1987; Maki & Serra, 1992), accessibility of information in memory (Baker & Dunlosky, 2006; Morris, 1990), ease of processing the text (Dunlosky & Rawson, 2005; Dunlosky, Rawson, & Hacker, 2002; Maki, Foley, Kajer, Thompson, & Willert, 1990; Rawson & Dunlosky, 2002), and global characteristics of texts such as length or difficulty (Weaver & Bryant, 1995). However, these cues may or may not lead to accurate judgments of test performance, depending on the nature of the test that is given. Some of these cues may produce good monitoring accuracy when tests are memory-based, but not when the tests require understanding of connections, or the generation or recognition of inferences based on the text. To understand the cues that may predict performance on these sorts of tests requires bridging theories of metacognitive monitoring with theories of comprehension (Rawson et al., 2000; Wiley et al., 2005; Weaver, 1990).

Several successful interventions have been informed by such an approach, which has been called a *situation model approach to metacomprehension* (Griffin, Wiley & Thiede, 2008; Thiede et al., in press; Wiley et al., 2005). This approach is based on the comprehension framework of Kintsch (1994, 1998) which posits that a reader creates multiple representations of a text as he or she reads. For instance, the reader constructs a representation of the text at a

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

surface level (e.g., the exact words), a textbase level (e.g., the meaning of sentences), and the situation-model level, where connections are made across units of the text as well as with prior knowledge. A well-constructed situation model integrates across the ideas contained in a text and allows the reader to form a causal model and inferences implied by the text. When tests of comprehension actually tap the situation model of a text (Kintsch, 1994; McNamara, Kintsch, Songer, & Kintsch, 1996; Wiley et al., 2005), metacomprehension accuracy should increase if readers use cues that tap the situation model level of representation to judge their comprehension. Furthermore, if readers are using cues other than those related to the situation model, monitoring attempts might be misdirected, which would result in poor metacomprehension accuracy.

Support for the situation model approach has been found across a number of studies. Thiede and colleagues (i.e., Thiede & Anderson, 2003; Thiede, Anderson, & Therriault, 2003) were able to increase monitoring accuracy from .27 to .60 with the use of summarizing and keyword listing tasks that were performed prior to judgment. However, this improvement only occurred when the tasks were performed at a delay after reading and not when performed immediately. The authors explained this delay effect in terms of whether the generation task involved accessing STM or LTM representations of the texts. However, subsequent work by Thiede, Dunlosky, Griffin, and Wiley (2005) demonstrated more convincingly that these previous effects were due to performing a generative task that required accessing and employing one's text representation after a delay. Readers are getting access to cues when they access their text representations and these cues are more predictive of comprehension test performance when accessed at a delay. This is true, even though the keyword listing task itself could be considered little more than a simple word recall task. Thiede et al. (2005) argued that the key factor is the

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

level of representation being accessed. When performed immediately, these generation tasks can be performed using the highly accessible surface representation, but at a delay the situation-model representation is more likely to be accessed due to the reduced accessibility of surface model. This interpretation is based on the work of Kintsch, Welsh, Schmalhofer and Zimny (1990; see also Fletcher & Chrysler, 1990; Schmalhofer & Glavanov, 1986), which has shown that access to surface information decays rapidly, whereas access to the situation model is more robust over time. As surface memory is less accessible at a delay, it is less likely to be used by readers as a basis for their comprehension judgments. Thus, the cues produced by the same task of recalling keywords varied in the degree of predictive validity when performed immediately versus at a delay, due the difference in the level of representation involved in performing the task.

At this point, it is important to clarify that cues which differ in their validity may not always differ in terms of the general cue type they represent. Cues can be categorized into different broad types, such as superficial (e.g., familiarity, interest), memory-based (e.g. recallability), and comprehension-based (e.g., ability to self-explain). Cues are valid when they happen to reflect the level of representation being assessed at testing. Certain cue types (e.g., ability to self-explain) may have consistent ties to a certain level of representation (e.g., the situation model). However, some cue types, like ability to recall keywords, may reflect varying levels of representation, depending on contextual factors like the time elapsed since reading. Thus, our discussions about readers using more valid cues may sometimes involve switching from one cue type to another, but may sometimes involve the same general cue type becoming a more valid predictor of performance.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Griffin et al. (2008) presented converging evidence for the situation-model approach by showing that directing readers toward their situation model via a "self explanation" instruction improved relative accuracy (r = .63). In this study, the self-explanation instruction prompted readers to explain the meaning and relevance of each part of the text to other parts and to the overall purpose of the text. Such explanation-based reading tasks have been shown to focus readers on their situation-model representations (Chi, 2000; Wiley & Voss, 1999). A further important point is that explanation occurred during reading and not at a delay. As no delay was involved, some alternative interpretations of the previously observed improvements in accuracy due to delayed generation effects (such as transfer-appropriate monitoring) are not viable explanations for the self-explanation effect.

In all these prior studies, the presumption is that the manipulations are improving access to the situation model, or more specifically, improving access to cues related to the quality of the situation model, and as a result, the interventions shift readers from monitoring poor cues to better cues for predicting their own comprehension. However, previous studies offer only indirect evidence to support this presumption. In Experiment 1, the delayed summary paradigm of Thiede and Anderson (2003) is again employed, but in addition, readers are asked to report the cues they are using to judge comprehension, thus providing the first direct investigation of this issue. Another extension to Thiede and Anderson's original study in the present investigation is in terms of the sample that was run.

The original study did not explore whether the delayed summary intervention might be effective for readers of all ability levels. Previous work has suggested that less-able readers might have poorer metacognitive skills than more-able readers (Garner, 1987). However, no studies have specifically investigated the metacomprehension accuracy of at-risk college readers.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Griffin et al (2008) recently showed that variability in reading comprehension skill among normal college readers was significantly related to metacomprehension accuracy. The present inclusion of at-risk readers allowed us to examine whether a robustly successful intervention like delayed generation serves to widen, narrow, or simply maintain the accuracy gap between more and less skilled readers. Also, if at-risk readers show both lower accuracy and reliance on more superficial judgments cues, then this would support the claim that cue validity is a critical factor in determining monitoring accuracy.

Thus, a main goal of the current studies was to explore whether we might observe a relation between ability and metacomprehension accuracy when comparing a typical college reader sample to a sample that the university required to attend remedial reading classes. Assuming such an effect would be found, of interest is exploring the possible reasons for poor performance among the at-risk reader reading sample, and in turn, what instructional contexts might address those issues and increase metacomprehension accuracy among at-risk readers.

Experiment 1

The primary purpose of this experiment was to further test the situation model approach to improving metacomprehension accuracy by evaluating whether the use of cues relevant to the situation model is associated with higher levels of accuracy. In this experiment, we replicate the procedures of Thiede and Anderson (2003) with both a typical college sample as well as a sample of college students who were required to enroll in remedial reading courses by the university. Students from both samples completed all three conditions: no summary, immediate summary and delayed summary. Each condition was run in a separate session, and order of conditions was counter-balanced. In addition, students were asked to report the basis for their judgments of comprehension.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

This is a more direct way of ascertaining cue use than has been used in previous research, which has relied on correlational data and the effects of targeted manipulations to infer shifts in the bases of metacomprehension judgments. Thus, in Experiment 1, the effects of different summary conditions were tested on both normal and at-risk college reader samples. And, the relation between cue use and metacomprehension accuracy was analyzed to explore possible explanations for poor accuracy, as well as potential differences between the two reader groups

Method

Participants. One hundred forty-two college students participated as a course requirement. Of the 142 who began the study; 15 (10.5%) failed to complete all three sessions and were dropped from the study. Although 127 students completed each of three sessions required for this experiment, 21 participants failed to respond to open-ended questions about cue use or had indeterminate gamma correlations, due to invariance in metacomprehension judgments; thus, only 106 had complete data. Of these, 32 were students recruited from a developmental reading course. These students (who had a mean ACT score of 14.2) were classified as at-risk readers by the university on the basis of their ACT scores (<18), and were required to enroll in the remedial reading course. In addition, 74 were students recruited from an introductory psychology course (who had a mean ACT score of 23.2) who were not required to take a remedial reading course. Although ACT score is not a pure test of reading ability, this was the measure used by the university to assign students to the remedial reading course, so it is used as the reading ability criterion for this study. The samples will be referred to as

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

"typical college" and "at risk" readers in reflection of the manner in which they were selected.

Although at-risk readers were actively recruited, the pool of students enrolled in the developmental reading course was far fewer than that enrolled in the psychology course; thus, it was not possible to obtain equal numbers of readers from the two groups in the study. All participants were treated in a manner consistent with the ethical standards of the American Psychological Association.

Materials. The texts were adapted from ACT test preparation materials. They ranged in length from approximately 600 to 800 words, and had an average Flesch-Kincaid readability score of 11.4. Three sets of five texts were constructed with a balance of topics from three general categories: natural science, social science, and humanities. The tests contained 10 multiple-choice items (with four alternatives) designed to assess comprehension (inference-making or application), rather than memory of details contained in the text (an example is presented in the Appendix).

Design. This study utilized a mixed design with reading group being a between-subjects variable and summarization condition being a within-subjects variable. That is, each participant completed each of summarization conditions in a separate session: no-summary, immediate summary, and delayed summary. Experimental sessions were separated on average by five days. A Latin-square design was used to counterbalance the order of tasks. Order was manipulated as a means of control, and was not expected to interact with the other variables of interest. A set of preliminary analyses confirmed that there were no significant order effects or interactions with outcome variables, all Fs < 1.9, $p > .10$. Hence, order is not considered in the main experimental analyses.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Procedure. All participants were instructed that they would read texts on a computer screen, rate their comprehension for each text, and then answer test questions for each text. They were also instructed that they might be asked to write a summary for some of the texts. Finally, they were instructed that they would respond to some questions regarding the tasks in the experiment.

Following instructions for the first session, the participants read a sample text and rated their comprehension of the text. The comprehension rating was prompted with the title of the text at the top of the computer screen and the question (as in Glenberg and Epstein, 1985), *"How well do you think you understood the passage whose title is listed above? 1 (very poorly) to 7 (very well)."* After rating their comprehension, participants answered a practice cue report question, which asked, *"When you finish reading text material, how do you decide whether you have understood the passage? That is, when asked to "grade" your comprehension of that passage, what do you base your grade on so you can say, 'I understood this passage well' or 'I read it, but I didn't understand it'?"* After typing their response to the practice question, they answered sample test questions.

During the no-summary task, participants read five texts. After reading all texts, they rated their comprehension for each of the texts. (For comparison to other studies, this represents the standard *delayed judgment* condition, because judgments are made after all texts are read and not immediately after reading.) After rating their comprehension of the last text, participants responded to two open-ended questions. One question required a *global* response and asked, *"You just rated your comprehension of five passages. What did you use to decide whether your comprehension over a passage*

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

was given a high rating or a low rating?" The computer then showed them the title of a passage they rated low and the title of a passage they rated high, and required a *comparative* response, *"You gave a lower rating to your comprehension for the passage entitled* (Title A) *and a higher rating to the passage entitled* (Title B). *Describe the differences between the passages and your reading experience that made you give different ratings of comprehension."* The participants then answered the 10 multiple-choice questions for each text.

During the immediate-summary task, participants read the first text displayed on the screen. They were then shown the title of the text and instructed to write a summary of that text. Once they finished writing the summary, they were presented with the next text. They read and immediately wrote a summary of each of the five texts. After writing the summary of the last text, participants rated their comprehension of each text and then answered the *global* and *comparative* cue-use questions. After typing their responses to the questions, they answered the 10 multiple-choice test questions for each text.

During the delayed-summary task, participants read all five texts. They were then shown the title of the first text they had read and were instructed to write a summary of that text. When they were finished with this summary, they were presented with the next title and asked to write a summary of that text, and so on for all texts. After writing a summary of the last text, participants judged their comprehension of each text and answered the *global* and *comparative* cue-use questions. After typing their responses to the interview questions, participants answered the 10 multiple-choice test questions for each text.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

For all conditions, the texts were presented in a randomized order for each participant. Texts were rated for comprehension and tested in the same order as they were presented for reading. After answering the last multiple-choice test question in each task, participants were presented with the number of questions they had correctly answered over all five tests. That is, they received feedback regarding overall performance; they did not receive feedback regarding performance on a test for a particular text. They then responded to a closed-ended *test-expectation* question which asked, *"Were you surprised at the score you got on the comprehension questions?"* For the immediate-summary and delayed-summary conditions, the participants also responded to a closed-ended *summary-use* question which asked, *"Did you think about your summary when you made your rating for comprehension?"*

Coding. Responses to the open-ended questions provided self-report information on the different cues used to judge comprehension. For the *global* and *comparative* cue use questions, a research assistant, who was blind to the condition, compiled a list of 30 cues that participants reported using to judge comprehension. These cues were collapsed into five cue types: ability to explain meaning (e.g., "I gave it a high number if I thought I could explain the meaning of the story to another person."), ability to recall or restate information about the text (e.g., "I based my rating on how well I could remember the ideas of the article."), prior knowledge of a topic (e.g., "I gave it a high rating because I knew a lot about the topic."), interest in the topic of the text (e.g., "I gave it a low rating because I think Beethoven is boring."), and use of features of the text including difficulty, ease of processing, readability, length and specific vocabulary (e.g., "I gave it a low rating because it was long and hard to read."). Readers' responses were not restricted and

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

could represent more than one of these 5 cue types. A second research assistant coded approximately 30% of the responses. The inter-rater reliability was quite high (kappa = .93).

Actual cue use is not directly observable to researchers, so any measure of cue use will have potential limitations. Self-reports of cognitive processes are a general concern (as described by Nisbett & Wilson, 1977). Readers may not have access to or awareness of the cues they rely upon, and they could just randomly report any plausible cue types that come to mind. Alternatively, readers could be generally biased towards reporting cues that seem more sophisticated than what they actually using. Such measurement errors are largely a problem for conclusions about the absolute levels on univariate distributions. However, these potential measurement problems are manifested as null-results in multivariate analyses, so the self-reports can be validated via their systematic relationships with other measures (see Erricsson & Simon, 1980). The primary focus of the present studies will be the multivariate relationship between reported cue use and metacomprehension accuracy, and how these covary across different reading levels and experimental conditions.

For the *test-expectation* question, participants overwhelmingly (over 90%) responded that the test was what they expected and that they were not surprised by the kind of test questions they had received. As a result, this question yielded no useful information for the purposes of this study and is not discussed further.

Results

The first step in analysis was to see whether the effects in the delayed summary condition replicated the earlier work. Metacomprehension accuracy was operationalized

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

as the gamma correlation between comprehension ratings and performance on a test of reading comprehension computed across texts following the procedure of Thiede and Anderson (2003; see Nelson, 1984, for a rationale for using gamma). Thus, before computing metacomprehension accuracy, descriptive analyses on judgments and test performance are presented. Then metacomprehension accuracy is considered. This is followed by an analysis of which cues were reported to be used as a basis for metacomprehension judgments and how cue use related to accuracy.

Metacomprehension judgments. The median of metacomprehension judgments across the five texts was computed for each participant. The median is the recommended measure of central tendency for small sets of scores where extreme scores could affect the mean (Gravetter & Wallnau, 1999). The mean of the medians was computed across participants by condition.

A 2 (reading group) x 3 (summary condition) ANOVA showed there was a main effect for reading group, F (1, 104) = 8.14, MSE = 1.03, $p < .01$, $\eta^2 = .07$. As seen in Table 1, judgments were higher for typical readers than at-risk readers. There was also a main effect for summary condition, F (2, 208) = 4.07, MSE = .75, $p < .02$, $\eta^2 = .04$, with follow-up tests indicating higher judgments in the no summary condition, compared to the other two. The interaction was marginal, F (2, 208) = 2.26, MSE = .75, $p < .10$, $\eta^2 = .02$. Importantly, similar variance in judgments was seen across conditions and reading groups, and there were no ceiling or floor effects.

Test Performance. The median proportion of correct test responses across the five texts was computed for each participant. The mean of the medians (presented in Table 1) was then computed across participants within each condition.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

A 2 (reading group) x 3 (summary condition) ANOVA revealed a significant main effect for reading group, F (1, 104) = 46.2, MSE = .02, $p < .0001$, $\eta^2 = .31$. The main effect for summary condition was marginal, F (2, 208) = 2.24, MSE = .01, $p < .10$, $\eta^2 = .02$. The interaction was significant, F (2, 208) = 3.40, MSE = .01, $p < .04$, $\eta^2 = .03$. Follow-up tests revealed that typical college readers performed better on these ACT-type passages than the at-risk readers, as would be expected since the samples were selected based on actual ACT scores. Moreover, the interaction was due to the typical readers doing worse in the immediate summary condition than in the other two conditions. Importantly, both groups showed similar variance in their performance and there were no ceiling or floor effects.

Metacomprehension Accuracy. Metacomprehension accuracy was operationalized as the gamma correlation between comprehension ratings and test performance across a set of texts. In this study, three intra-individual correlations were computed for each participant, one for each summarization condition. The mean gamma correlation was then computed across participants for each condition. As seen in Figure 1, metacomprehension accuracy differed significantly across conditions, F(2, 208) = 19.3, MSE = .22, $p < .001$, $\eta^2 = .16$. Consistent with the findings of Thiede and Anderson (2003), follow-up tests found that the delayed-summary condition increased accuracy over the immediate-summary and no-summary conditions, which did not differ.

A main effect of reading group was also found, as metacomprehension accuracy was greater for typical readers than for at-risk readers, F(1, 104) = 4.97, MSE = .09, p = .03, $\eta^2 = .05$. Summary condition did not interact with reading group, $F < 1$. The lack of an interaction indicates that the delayed summarization instruction was not a strong

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

enough intervention to equate the accuracy of the two reading groups. As a result of the two main effects, typical readers reached the highest level of accuracy in the delayed-summary condition (around .60) and remained more accurate than the at-risk readers who only achieved accuracy (around .45) even in the delayed-summary condition.

Cues used to Judge Comprehension

The responses to the *global* and *comparative* cue use prompts revealed largely similar distributions of cue use across conditions and reading groups. Because of their extreme similarity, only data for the *global* prompt are presented here. As mentioned above, comments were originally sorted into five categories. The frequency of responses from this initial coding is presented in Table 2. Note that in this table, participants can contribute to more than one cue type in each condition.

In order to create an exclusive coding system for analyses, readers were classified into one of four cue-use profiles: *surface, reader, memory, or comprehension*. Readers who reported using any cues related to the qualities of the text itself were classified as fitting the *surface profile,* regardless of any other cues that were reported. Those who reported relying on their own ability to understand or explain the text, but not surface cues, were classified as using *comprehension-based cues*. Readers who referred to their ability to recall the text, but not comprehension- or surface-based cues were classified as using *memory-based cues*. Finally, readers who reported relying on judgments about their own level of familiarity with or interest in the topic, without mentioning the text's surface features, memory-based cues, or comprehension-based cues were classified as relying on *reader-based cues*. Because of low numbers of observations in the prior

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

knowledge and interest categories, these two cue sets were collapsed into a single "reader characteristics" category.

The decision to assign all readers using *any* surface cues to a surface profile was driven by observation of the data in the neutral condition. First, looking at the readers who reported only surface cues, we observed very low gammas among this group. The participants who reported using only surface cues had a mean gamma of -.03, while participants who reported only reader-based cues had a mean gamma of .19, those reporting only memory-based cues had a mean gamma of .20, and those reporting only comprehension-based cues had a mean gamma of .71.

Next we examined the performance of participants who reported a combination of cue types from multiple categories. For both surface/reader combinations (-.13), and surface/comprehension combinations (-.33), the gammas were quite dissimilar from those for readers who used exclusively reader and comprehension profiles. (Although for readers in the memory-based profile, the combination with surface cues if anything improved performance, .35). Given that in most cases, reporting use of any surface cues made readers appear more similar to those who reported only surface cues, and the point of this analysis was to attempt to characterize the behaviors that related to accurate or inaccurate metacomprehension, we elected to collapse all the combinations that included surface cues into the surface profile category. By the same logic, we examined the effects of reader cues in combination with memory (.36) and comprehension-based (.84) cues. In both cases, behaviors were consistent with performance in the pure "memory-based" and "comprehension-based" conditions. Thus in these cases the combinations

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

were collapsed into the higher-order categories. By this process, we determined our classification of cue use profiles to be used for all conditions.

The above classification was done three times for each individual, once for each summary condition based on their comments at the end of each condition. The overall frequency of readers falling into each profile type by reading group and summary condition are presented in Figure 2. First note that, overall, comprehension-based profiles were the least common, while memory-based profiles were the most common (i.e. cue use related to the ability to recall information from the text). Second, note that almost half of at-risk readers had a surface-cue profile, whereas typical readers were most likely to fall into the memory-based profile.

Further, the distribution of profiles across conditions changed especially for the typical readers – who focused less on reader characteristics, and more on the quality of their ability to recall a text, when they made judgments in the delayed-summary condition. Pairwise Wilcoxon tests revealed no differences in distributions across summary conditions for at-risk readers (Zs < .57), whereas the distribution in the delayed-summary condition was different than the distribution for the other two conditions (Z=2.45 and Z=2.83, ps <.01) among the typical readers, with no difference between immediate-summary and no-summary conditions (Z < .54).

Relation between cue use profile and metacomprehension accuracy. Several analyses were performed to explore the effect of summary condition on metacomprehension accuracy and whether cue use was related to accuracy. First, within each of the summarization conditions, between-subjects analyses were conducted to examine differences in metacomprehension accuracy due to cue use profiles. Next,

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

within-subjects analyses were performed on subsets of participants who fit the same profile across conditions. Finally, a *best-cue* analysis was used to create a stable within-subjects variable related to cue use, so that a fully within-subjects model could be tested.

Within Summary Condition, Between Subjects Analysis. The overall patterns of metacomprehension accuracy for cue use and summary condition are presented in Figure 3. Note that in this figure, each participant has a monitoring accuracy score for each summary condition, but the cue use profile that an individual is assigned to can change across conditions. Thus, to analyze these data, we performed a separate ANOVA for each summary condition.

For the no-summary condition, a 2 (reading group) x 4 (cue use profile) ANOVA revealed a main effect for cue use profile, $F(3,101)=2.70$, $MSE=.30$, $p<.05$, $\eta^2 = .08$, but no effect for reading group, $F < 1$. (The interaction is not reported due to a lack of data in the at-risk reader/comprehension cell.) Follow-up tests revealed that monitoring accuracy was significantly worse for readers who fit a surface cue profile, and accuracy was significantly better for readers who fit a comprehension-based profile, than for other profiles. Accuracy for reader-based and memory-based profiles did not differ.

For the immediate-summary condition, a 2 (reading group) x 4 (cue use profile) ANOVA revealed a main effect for cue use profile, $F(3,101)=3.02$, $MSE=.22$, $p<.03$, $\eta^2 = .08$, but no effect for reading group, $F < 1$. (The interaction is not reported due to a lack of data in the at-risk reader/comprehension cell.) Follow-up tests revealed that monitoring accuracy was significantly better for readers who fit a comprehension-based profile than for all other profiles. Accuracy for the remaining profiles did not differ.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

For the delayed-summary condition, a 2 (reading group) x 4 (cue use profile) ANOVA revealed a main effect for cue use profile, $F(3,98)=11.9$, $MSE=.11$, $p<.0001$, $\eta^2 = .27$, but no effect for reading group and no interaction, Fs < 1. Follow-up tests revealed that monitoring accuracy was significantly worse for readers who fit a surface-cue profile, and accuracy was significantly better for readers who fit a comprehension-based profile, than for other profiles. Accuracy for reader-based and memory-based profiles did not differ.

Consistent Cue Use Profile, Within-Subjects Analysis. The analyses within each summary condition revealed that readers who rely on cues based on surface features of the text had lower metacomprehension accuracy, and those who rely on comprehension-based cues had greater accuracy. However, an interesting pattern can also be seen if one looks across summary conditions, as it appears that the utility of using memory-based and reader-based cues changes, and that such cues are only predictive in the delayed-summary condition. To test that increases in predictive accuracy are due to the summary condition, and not due to particular individuals who only fall into a memory-based profile in one condition but not the other, we computed the average gammas for only the subset of participants who fell into the *memory-based* profile in both immediate- and delayed-summary conditions (N=30). For these participants, gammas were significantly higher in the delayed-summary condition (.67, *SE* .06) than in the immediate-summary condition (.25, *SE* .07), $t(29)= 4.9$, $p<.001$. Thus, relying on memory-based cues as a basis for predictive judgments can be a particularly effective strategy, but this is only the case when these judgments follow summaries that are generated at a delay.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

When the same analysis is performed for the *reader-based* profile, only 4 participants fell into this category in both the immediate- and delayed-summary conditions, and their means did not differ (immediate: $M = .45$, SE .20; delayed: $M = .44$, SE .28, $t<1$). Thus the change between relatively low metacomprehension accuracy in the no-summary and immediate-summary conditions, and high accuracy in the delayed-summary condition, is due to movement of individuals into different profile types across conditions.

Best Cue Reported, Within Subjects Analysis. In order to compare across summary conditions in a more powerful way using the full sample, a further analysis assigned individuals to a single cue basis category, as a function of the highest quality cue that was used in any of the three summary conditions.

The order of cue quality was based on theoretical premises that metacomprehension cues that assess the quality of understanding of the situation model level representation will be the most valid predictors of performance on a test of comprehension (Griffin et al., 2008; Thiede et al., in press; Wiley et al., 2005). Therefore, reports of cues related to the quality of understanding or ability to explain the content of the passage were rated as highest in quality (i.e. the comprehension-based cue category described above).

The next set of cues in terms of quality were the memory-based cues that again referenced a readers' reflection on their own representation, but commented specifically on the ability to remember the text (as opposed to the ability to understand it). This level of monitoring activity corresponds theoretically to assessing the quality of the textbase, which can be a predictor of performance on comprehension tests in some cases, but this is

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

not necessarily true and may be predictive to a lesser degree than situation-model level judgments (Wiley et al., 2005).

The final two sets of cues were classed as lower in quality, because neither required readers to reflect on their own representation of the texts. The third class of comments was those that referred to predictions based in reader characteristics of personal interest or familiarity with the content of the texts. These cues can be predictive–having no interest in a test may accurately predict very low performance on a subsequent test due to a lack of motivation–but importantly they do not require reflection on or access to one's own internal representation of the text. These comments instead refer to a quality of the reader, so they are self-assessments, but they may not necessarily relate to the comprehension of a particular text.

The lowest class of cues was those that referred to qualities of the text itself–mainly the readability of text, the difficulty of the vocabulary used, and length of the passage. Again, these cues can be predictive of test performance, but they are heuristic approaches that do not require reflection on internal representations. As with all heuristics, they may lead to predictive judgments in some cases, but especially when comprehension performance is being predicted, they may be misleading.

Thus, each participant was assigned to a single level of cue use based on the *highest quality* cue that was ever reported by the individual in any condition. This single measure of the *best cue reported* allowed for comparisons across summary instructions because with this coding individuals do not contribute to different categories across conditions. Therefore this approach provides an additional way to assess the effects of delayed summaries on metacognitive accuracy.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Best Cue Reported as a function of Reading Group. Splitting the cue quality into heuristic (surface, reader) and monitoring (memory-based, comprehension-based) categories, we found that monitoring cues were more frequently used by typical college readers than for at-risk readers (87.1% vs. 66.6%), while heuristic cues were more frequently used by at-risk college readers than typical readers (33.3% vs. 12.9%).. This resulted in a significant chi-square $\chi^2(1)=6.27$, $p< .01$, showing that distribution across the two best cue categories differed by reading group.

Metacognitive Accuracy by Best Cue Reported. Figure 4 presents the average metacomprehension accuracy for each *Best Cue* group as a function of summary condition. A 3 (Summary Condition) x 4 (Best Cue group) ANOVA revealed significant effects for both summary condition, $F(2, 204) = 12.3$, $MSE=.21$, $p<.0001$, $\eta^2 = .11$, and best cue group ($(3, 102) = 7.03$, $MSE=.08$, $p<.001$, $\eta^2 = .17$). (There was no main effect for Reading Group, $F<1$, once cue use was included, so it was not included in the model.) The interaction between Summary Condition and Best Cue group was also significant, $F(6, 204) = 2.57$, $MSE=.21$, $p<.02$, $\eta^2 = .07$.

Follow-up tests for the summary condition effect revealed that metacomprehension accuracy in the delayed-summary condition was better than in the immediate-summary condition, which in turn was better than in the no-summary condition.

Follow-up tests for the *best cue* effect revealed that use of comprehension-based cues led to better accuracy than all other cues. Use of memory-based cues was significantly worse than comprehension-based cues, but significantly better than surface

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

or reader-based cues. Accuracy among those reporting use of reader and surface cues did not differ.

To follow up the significant interaction between summary condition and best cue group, we tested for the presence of a main effect for summary condition within each cue condition. Within cue condition ANOVAs revealed that accuracy of readers whose best cue was in the surface or reader categories did not change as a function of summary condition. However, accuracy for readers whose best cue was memory-based did improve specifically in the delayed-summary condition. Further, accuracy for readers whose best cue was comprehension-based improved in the immediate-summary condition over the no-summary condition, and improved again with the delayed-summary condition over the immediate-summary condition, thus resulting in the highest level of metacomprehension accuracy in our sample.

Summary from Cue Analyses. The findings from the *best cue* analysis converge with other cue-use analyses. Comprehension-based cues were the best predictors of performance on comprehension tests, but were rarely used.

Memory-based cues were able to lead to valid predictions of test performance, but interestingly, only in the delayed-summary condition. This shift in the validity of memory-based cues is consistent with the explanation of the delayed summary effect as being a function of the changes in memory for text that occur after a delay (Thiede et al., 2005). Over time, memory for surface information fades while situation-model level information remains. Thus, when readers base their cues on their ability to remember a text, which become apparent during a summarization task, their judgments will be more

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

predictive of comprehension test performance as long as some time passes after reading but before attempting to summarize. The present data provide support for this account.

Finally, judgments based on surface characteristics of the text were the least predictive of performance on comprehension tests. And, in general, at-risk readers were more likely to use surface and reader cues, and less likely than typical readers to engage in metacognitive processes of reflecting on either their own level of understanding or their ability to remember texts in order to generate their judgments. This finding may perhaps be due to resource limitations that make the process of constructing a representation of text, *and* reflecting on it, too demanding (Griffin et al., 2008). Thus, at-risk readers may have been forced to resort to heuristic approaches to guide their judgment process. This hypothesis led to the formulation of a specific goal for the second experiment: to provide a context for at-risk readers that may give them direct access to valid cues for judgment and, at the same time, might allow them to reflect on the quality of their representations of texts.

Use of Cues based in Summary Writing Experience. All of the above *cue use* analyses used the responses to the *global* question as a basis for determining the kinds of cues that readers were using. In addition, a final closed-ended *summary-use* question asked participants directly whether they thought about their summaries while making comprehension judgments. A large proportion of readers endorsed using this cue. Using information gained from the experience of writing a summary would be seen as an effective cue for judging comprehension. When readers have difficulty summarizing a text, this should alert them that their understanding is poor (Dunlosky & Rawson, 2005; Thiede & Anderson, 2003). In the immediate-summary condition, the proportion of

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

typical readers who reported thinking about their summaries (66%) did not differ from that of at-risk readers (61%), $\chi^2(1) = .01$. However, in the delayed-summary condition, a greater proportion of typical college readers reported thinking about their summaries as they made their judgments (78%) than did at-risk college readers (44%), $\chi^2(1) = 6.98$. Further, a 2 (reading group) x 2 (used summaries versus did not) ANOVA revealed that reported *use of summaries* affected accuracy in the delayed-summary condition. The main effect for summary use was significant, $F(1, 102) = 13.8$. Neither the main effect for reading group nor the interaction were significant, Fs < 1. For both reading groups, accuracy was greater for those who reported using summaries as a cue for judging comprehension (Typical Readers: $M = .73$, SE .05; At-Risk Readers: $M = .70$, SE .11) than for those who did not (Typical Readers: $M = .37$, SE .09; At-Risk Readers: $M = .44$, SE .10).

A 2 (reading group) x 2 (used summaries versus did not) ANOVA revealed that reported *use of summaries* did not affect accuracy in the immediate-summary condition. Neither main effect nor the interaction was significant, Fs < 1.

These data provide additional evidence that the cues provided by the experience of generating a summary were more predictive of comprehension performance in the delayed-summary condition. They also converge with previous analyses showing that metacomprehension accuracy varies as a function of the cues that are used as a basis for comprehension judgments. At-risk readers who reported using summaries as a basis for their comprehension judgments were just as accurate as the typical college readers.

Discussion

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Several important findings emerged from this study. First, the effectiveness of a delayed summary instruction was replicated in typical college readers and extended to a population of at-risk readers. Although the intervention did not close the gap between ability levels, it did improve metacomprehension accuracy overall. Second, based on self-reports of cues used as the basis for comprehension judgments, it appears that the benefits of the delayed-summary condition are indeed because it makes judgments based on memory-related cues more valid. As memory-based cues are the default basis for judgments for many readers, the delayed-summary condition improves metacomprehension accuracy by putting readers in a context where memory-based cues are predictive of comprehension performance. Across conditions, at-risk readers have less accurate judgments and were more likely to report using surface type cues when making those judgments. However, when at-risk readers did report using valid cues as the basis for their judgments (thinking about their ability to generate a summary), or not using poorer cues, then they were just as accurate as typical college readers. The fact that the reported cues can account for when at-risk and typical readers differ in monitoring accuracy lends support to the validity of this self-report measure of cue use. These findings are also consistent with the cue-utilization perspective on monitoring accuracy (Koriat, 1997).

Besides the increased validity of recall cues in the delay condition, there was evidence that the different conditions had some effect on the type of cues readers used. Especially in the delayed-summary condition, metacomprehension accuracy improved when readers used valid cues such as the ability to generate a summary or explanation of a text as the basis for their judgments. On the other hand, focusing on simple surface and

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

reader cues led to poorer accuracy. Importantly, this is the first study to attempt to provide direct evidence of the kinds of cues that readers use to judge comprehension. Perhaps most striking is how few participants spontaneously reported using cues that would be highly diagnostic of the quality of their situation models, i.e. the ability to explain the text. Only 11 readers mentioned this as a basis for their judgments. As disturbing is how many participants spontaneously reported surface features of the texts as the basis of their judgments of comprehension. This speaks to the need to give students a better understanding of what it means to comprehend expository text, so that they might base their comprehension judgments on more predictive cues (Wiley et al., 2005).

The differences in metacomprehension accuracy and cue use as a function of reading proficiency highlight the need to explore additional interventions. Although some at-risk readers were able to perform as well as typical readers, this depended on them selecting valid cues for their judgments. There were still a large number of at-risk readers who were focused on incorrect cues for comprehension. Thus, Experiment 2 explored an intervention that explicitly directed less-able readers toward appropriate cues for judging their comprehension.

Experiment 2

The results of Experiment 1 suggest that metacomprehension accuracy is influenced by the cues participants use to judge comprehension and that the validity of cues changed from one situation to another. Moreover, these results suggest that metacomprehension accuracy for many at-risk readers is compromised by the use of inappropriate cues (based on surface features of a text). In this experiment, we attempted

to change the cues used by less-able readers to judge comprehension. In particular, we attempted to direct their attention to cues related to the situation model of texts by instructing them to construct concept maps as they read the texts.

A concept map is a graphic representation of the underlying structure of the meaning of a text. Constructing concept maps can be an effective organizational strategy, which helps readers formulate the connections among concepts in a text (Weinstein & Mayer, 1986). Concept mapping was chosen as an intervention as it has been suggested that such an approach may be particularly helpful and appropriate for less-able readers (Stensvold & Wilson, 1990; for a review and meta-analysis on effectiveness of concept maps with low-ability learners, see Nesbit and Adesope, 2006). Concept mapping shares many similarities with argumentation and self-explanation tasks (Weinstein & Meyer, 1986), but because it employs the construction of external, visual representations while readers have access to the texts, it may place fewer demands on the reader than other explanation tasks. Instructing at-risk readers to construct a concept map of a text during reading should not only help them identify important connections, and therefore help them construct a situation model for a text, but it should also increase the salience of the quality of that situation model level representation, which they can then use to judge their comprehension of a text. Thus, we hypothesize that metacomprehension accuracy will improve when at-risk readers construct concept maps during reading (versus when they do not).

Method

Participants. Twenty-one students enrolled in a developmental reading course participated in the experiment as part of the course requirements (none of these students

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

participated in Experiment 1). As in Experiment 1, all of these students had ACT scores less that 18 (M = 12.2, SE = .54) which required their enrollment in the remedial reading course. (All participants were treated in a manner consistent with the ethical standards of the American Psychological Association.

Materials. The texts were adapted from on-line materials offered as supplementary readings for a developmental reading textbook. They ranged in length from approximately 250 to 350 words, and had an average Flesch-Kincaid readability score of 10. We constructed three sets of five texts with a balance of topics from three general categories: natural science, social science, and humanities. The sets of texts were randomly assigned across condition for each participant. The tests contained five multiple-choice items designed to assess comprehension (inference making or application), rather than memory of details contained in the text, an example of the texts and tests are in the Appendix.

Design. This study utilized a within-subjects design. Each participant first completed immediate-judgment and delayed-judgment conditions, with the order of these conditions counterbalanced. These conditions were completed on separate days with one week between sessions. The order of conditions did not affect any of the outcome variables, ts < 1.3; therefore the order of these conditions was not included in subsequent analyses. All students then completed concept map training and ran in the concept map condition.

Procedure. All participants were instructed that they would read texts on a computer, judge how well they understood each text, and then take a test for each text. In the delayed-judgment task (the control condition used in Experiment 1 which is the standard in the metacomprehension literature, Maki, 1998b), participants read all five texts, then made metacomprehension judgments for the texts in a block, and then answered test questions for the

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

texts in a block. The prompt for metacomprehension judgments in this experiment was the same as in Experiment 1.

A second comparison condition was also run using immediate judgments. In the immediate-judgment condition, participants read a text and judged their comprehension of the text immediately after reading. After making their judgment, participants answered five multiple-choice questions on the text. They completed this procedure for all five texts. Maki (1998a) has shown that an immediate judgment condition can produce higher levels of metacomprehension accuracy than the standard delayed-judgment condition. But the main reason for including this condition was because the concept map condition used immediate judgments following the construction of each concept map, and this control condition was needed to provide a well-matched comparison.

After completing these two conditions, participants received eight 50-minute class periods of instruction and practice constructing concept maps. They then completed the concept map condition, and received a new set of texts in a concept map condition on the day after the final period of instruction. In the concept-map condition, for each text in the set, participants constructed a concept map while reading. After reading and constructing a concept map, participants made their metacomprehension judgment (without access to their concept maps), and then answered five multiple-choice questions on the text. Participants read and constructed concept maps, judged comprehension, and answered test questions for all five texts. Thus, the procedure was like the immediate-judgment condition except that participants constructed concept maps while reading.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Results and Discussion

Test performance and comprehension ratings. As metacomprehension accuracy describes the relations between comprehension ratings and performance on a test of reading comprehension, descriptive statistics of these variables are reported first. The median proportion of correct test responses and metacomprehension judgments across the five texts was computed for each participant. The mean of the medians was then computed across participants within each condition. Test performance differed across conditions, $F(2, 40) = 15.3$, $MSE = .44$, $p < .001$, $\eta^2 = .43$ (see Table 1). Follow-up tests showed that test performance was greater for the concept map condition than the immediate-judgment condition, $t(20) = 4.0$, $p < .001$; or the delayed-judgment condition, $t(20) = 5.6$, $p < .001$. Thus, constructing concept maps while reading improved comprehension, which is consistent with the literature.

As seen in Table 1, the magnitude of metacomprehension judgments did not differ across conditions, $F(2, 40) < 1$.

Metacomprehension accuracy. Metacomprehension was again operationalized as the Goodman-Kruskal gamma correlation between metacomprehension judgments and performance on a test of reading comprehension computed across texts. Three gamma correlations were computed for each participant, one for each condition. The mean intra-individual correlation was then computed across participants for each condition. One participant in each condition had an indeterminate gamma correlation due to invariance in judgments. Metacomprehension accuracy differed across conditions, $F(2, 36) = 3.6$, $MSE = .22$, $p < .05$, $\eta^2 = .17$ (see Figure 5). Follow-up tests showed that metacomprehension accuracy was greater for the concept map condition than for the immediate-judgment/test condition, $t(18) = 2.3$, $p < .05$; or the delayed-judgment/test

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

condition, $t(19) = 2.4$, $p < .05$. No differences were found between the two control conditions, suggesting that a delay before judging neither helped nor hurt metacomprehension accuracy.

The above finding suggests that participants used something about their experience with their concept maps as a basis for their metacomprehension judgments. To evaluate this possibility, we first coded the number of appropriate connections made between concepts within each concept map as a metric of the quality of the representation. A second research assistant then scored the concept maps of 6 participants (approximately 30% of the responses) and scored the number of connections. The inter-rater reliability on coding these connections was quite high (kappa = .94). For each participant, we then computed a gamma correlation between the number of connections and metacomprehension judgments across the texts. The mean intra-individual correlation between the number of connections and metacomprehension judgments was .32 (*SEM* = .13), which is significantly different from zero, $t(19) = 2.5$, $p < .05$. This suggests that participants used the quality of their concept maps as basis for their judgments of comprehension.

For this to help explain the improved metacomprehension accuracy, concept maps would need to be predictive of test performance. To evaluate this possibility, we computed a gamma correlation between the number of connections and test performance across the texts. The mean intra-individual correlation between the number of connections and test performance was .38 (*SEM* = .11), which is significantly different from zero, $t(20) = 3.4$, $p < .01$. Thus, the number of connections included in concept maps was predictive of test performance. The concept maps thus served as a mechanism that made the quality of the situation model for each text salient to the reader, and therefore provided a predictive basis for metacomprehension judgments.

General Discussion

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

The present studies offer several important findings. The first is that self-reports of cue use indicated that poorer metacomprehension accuracy can be seen as a function of inappropriate cue use. That is, when readers do use appropriate cues to judge comprehension, they make accurate judgments. The second important finding is at-risk readers tended use more inappropriate cues. This finding led to the suggestion that low-ability readers might need interventions that aid them in selecting valid cues. In response to this issue, the final important result was that introducing a concept mapping intervention to a sample of at-risk readers improved both their comprehension and their metacomprehension as they attempted to learn from texts. This suggests that less-able readers benefited from a task that guided their learning of expository text, by teaching them to attend to the connections that could be made within each text. Consistent with previous research (e.g., Nesbit & Adesope, 2006; Stensvold & Wilson, 1990), we found that concept maps were effective learning tools for less-able readers. In the current study, we go beyond previous results to show that such an intervention improved metacomprehension accuracy in less-able readers as well as learning outcomes. Using a concept map intervention, less-able readers were able to reach a level of metacomprehension accuracy (around .60) that was comparable to the best levels that have been achieved in the literature.

The present studies provide additional support for the situation model approach to improving metacomprehension. One important property of concept-mapping is that it gets readers to attend to the quality of the situation model that they are constructing. The fact that improvements in metacomprehension accuracy were seen as a result of this activity is consistent with other studies that have improved metacomprehension via interventions that make the quality of the situation model salient to readers, including generating keywords or summaries at a delay (Thiede & Anderson, 2003; Thiede et al., 2005), or generating self-explanations (Griffin et al.,

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

2008). However, all of these studies have been conducted with skilled or typical reader populations. The present study extends previous work into a new population that sorely needs support in expository text comprehension. In Experiment 1 it was especially at-risk readers who used inappropriate cues to judge their own comprehension. The positive effects of concept mapping tasks observed here, suggest it may be a promising alternative to self-explanation that may be especially appropriate for younger or less-able readers. Such readers may not be able to handle the additional load imposed by monitoring, self-explanation or delayed-generation activities without the supportive assistance of the external representations that concept mapping provides.

Consistent with previous research that argues for the importance of using the situation model as a basis for comprehension judgments, we found that metacomprehension accuracy was greater for participants who reported using their ability to explain the meaning of texts as a cue for judging comprehension (versus those who did not report using this cue). Moreover, given that constructing concept maps may help less-able readers formulate a situation model for a text and attend to their situation model when judging comprehension, the findings from Experiment 2 provide additional evidence that getting readers to focus on their situation model during reading will improve metacomprehension accuracy. Further research now needs to be done to illustrate how and when students can translate monitoring accuracy into effective regulation of their own study behaviors, including making better choices of what to read and reread while studying, which in turn, will ultimately improve learning from expository text.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

References

Baker, J. M. C., & Dunlosky, J. (2006). Does momentary accessibility influence metacomprehension judgments? The influence of study—judgment lags on accessibility effects. *Psychonomic Bulletin & Review, 13*, 60-65.

Benjamin, A. S., & Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127,* 55-68.

Chi, M. T. H., DeLeeuw, N., Chiu, M., & Lavancher, C. (1994). Elicit self-explanation improves understanding. *Cognitive Science, 18,* 439-477.

Dunlosky, J. & Lipko, A. R. (2007). Metacomprehension: A Brief History and How To Improve Its Accuracy. *Current Directions in Psychological Science, 16,* 228-232.Dunlosky, J. & Rawson, K. A. (2005). Why Does Rereading Improve Metacomprehension Accuracy? Evaluating the Levels-of-Disruption Hypothesis for the Rereading Effect. Discourse Processes, 40, 37-55.

Dunlosky, J. & Rawson, K. A., & Hacker, D. J. (2002). Metacomprehension of science text: Investigating the levels-of-disruption hypothesis. In Otero, J., Leon, J. A., & Graesser, A. C. (Eds.). *The Psychology of Science Text Comprehension*. (pp. 255-279). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Dunlosky, J. and Thiede, K.W. (1998). What makes people study more? An evaluation of four factors that affect people's self-paced study. *Acta Psychologica, 98*, 37-56.

Ericsson, K. A. & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215 – 251.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Garner, R. (1987). *Metacognition and reading comprehension.* Norwood, NJ: Ablex.

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition, 15,* 84-93.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General, 116,* 119-136.

Goodman, L.A. & Kruskal, W. H. (1954). Measures of association for cross classification. *Journal of the American Statistical Association, 49*, 732-764.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48,* 163-189.

Gravetter, F. J. & Wallnau, L. B. (1999). *Essentials of Statistics for the Behavioral Sciences* (3rd Ed.). Pacific Grove, CA: Brooks/Cole Publishing Company.

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.

Kintsch, W. (1994). Learning from text. *American Psychologist, 49*, 294-303.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

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.

Maki, R.H. (1998a). Predicting performance on text: Delayed versus immediate predictions and tests. *Memory & Cognition, 26*, 959-964.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Maki, R. H. (1998b). Test predictions over text material. In Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). *Metacognition in Educational Theory and Practice.* (pp. 117-144). Hillsdale, NJ: LEA.

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*, 609-616.

Maki, R. H., Jonas, D. & Kallod, M. (1994). The relationship between comprehension and metacomprehension ability. *Psychonomic Bulletin & Review*, 1, 126-129.

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.

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, *97*, 723-731.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43.

Nisbett, R. E. & Wilson. T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General, 131,* 349-363.

Morris, C. C. (1990). Retrieval processes underlying confidence in comprehension judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 223-232.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp. DOI: 10.1080/01638530902959927

Nelson, T. O. (1984). A comparison of current measures of feeling-of-knowing accuracy. *Psychological Bulletin, 95*, 109-133.

Nelson, T.O. and Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G.H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125-141). New York: Academic Press.

Nesbit, J. C. & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research, 76*, 413-448.

Nist, S.L. & Holschuh, J.L. (2000). Comprehension strategies at the college level. In R.C. Flippo & D.C. Caverly (Eds.), *Handbook of college reading and study strategy research* (pp. 75-86). New York: Academic Press, Inc.

Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 69-80.

Rawson, K., & Dunlosky, J., & Thiede, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition, 28*, 1004-1010.

Rinehart, S. D. & Platt, J. M. (2003). Metacognitive awareness and monitoring in adult and college readers. *Forum for Reading, 15,* 54-62.

Royer, J. M., Carlo, M. S., Dufrense, R., & Mestre, J. (1996). The assessment of levels of domain expertise while reading. *Cognition and Instruction, 14,* 373-408.

Stensvold, M. S. & Wilson, J. T. (1990). The interaction of verbal ability with concept mapping in learning from a chemistry laboratory activity. *Science Education, 74,* 473-480.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Thiede, K. W. & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology, 28,* 129-160.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95,* 66-73.

Thiede, K. W., Dunlosky, J., Griffin, T. D., & Wiley, J. (2005) Understanding the delayed keyword effect on metacomprehension accuracy. *Journal of Experiment Psychology: Learning, Memory & Cognition*, *31*, 1267-1280

Wagoner, S.A. (1983). Comprehension monitoring: What it is and what we know about it. *Reading Research Quarterly, 18,* 328-346.

Weaver, C. A. (1990). Constraining factors in calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 214-222.

Weaver, C. A., III, & Bryant, D. S. (1995). Monitoring of comprehension: The role of text difficulty in metamemory for narrative and expository text. *Memory & Cognition, 23,* 12-22.

Weinstein, C.E. & Mayer, R. E. (1986). The teaching of learning strategies. In M. C. Wittrock (Ed.), *Handbook on Research in Teaching* (3rd ed., pp. 315-327). New York: Macmillan

Weinstein, C.E. & Rogers, B.T. (1985). Comprehension monitoring: The neglected learning strategy. *Journal of Developmental Education, 9,* 6-9, 28-29.

Wiley, J. (2001) Supporting understanding through task and browser design. *Proceedings of the Twenty-third annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

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

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.

Author Notes

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305H03170 and R305B07460 to Keith Thiede and Jennifer Wiley. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Footnotes

1. 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 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).

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Table 1

Descriptive Statistics on Metacomprehension Judgments and Test Performance by Condition, Reading Group and Experiment

Conditions	Judgments	Test Performance
Experiment 1—At-RiskReaders		
No Summary	4.28 (.19)	4.20 (.03)
Immediate Summary	3.71 (.24)	4.22 (.03)
Delayed Summary	3.94 (.22)	4.10 (.03)
Experiment 1—Typical Readers		
No Summary	4.72 (.13)	6.45 (.02)
Immediate Summary	4.65 (.16)	5.83 (.02)
Delayed Summary	4.41 (.15)	6.28 (.02)
Experiment 2—At-Risk Readers		
Concept map	4.57 (.29)	3.71 (.14)
Immediate judgment/test	4.67 (.28)	2.81 (.18)
Delayed judgment/test	4.62 (.24)	2.67 (.14)

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

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Table 2

Number (Proportion) of Participants who Reported Basing Comprehension Ratings on a Particular Cue by Condition and Reading Group in Response to Global Cue Use Question

Cues Reported	At-Risk Readers	Typical Readers
No Summary		
Surface features	12 (.33)	15 (.21)
Prior knowledge	12 (.33)	22 (.31)
Interest	17 (.47)	30 (.43)
Memory	15 (.42)	33 (.47)
Comprehension	0 (.00)	4 (.06)
Immediate Summary		
Surface features	14 (.39)	17 (.24)
Prior knowledge	10 (.28)	23 (.33)
Interest	12 (.33)	33 (.47)
Memory	18 (.50)	36 (.51)
Comprehension	0 (.00)	7 (.10)

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

	Delayed Summary	
Surface features	15 (.42)	11 (.16)
Prior knowledge	17 (.47)	21 (.30)
Interest	13 (.36)	18 (.26)
Memory	18 (.50)	49 (.70)
Comprehension	1 (.03)	10 (.14)

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

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Figure Captions

Figure 1. Mean metacomprehension accuracy by summary condition and reading group. The error bars represent the standard error of the mean.

Figure 2. Proportion of participants in each summary condition by cue use profile.

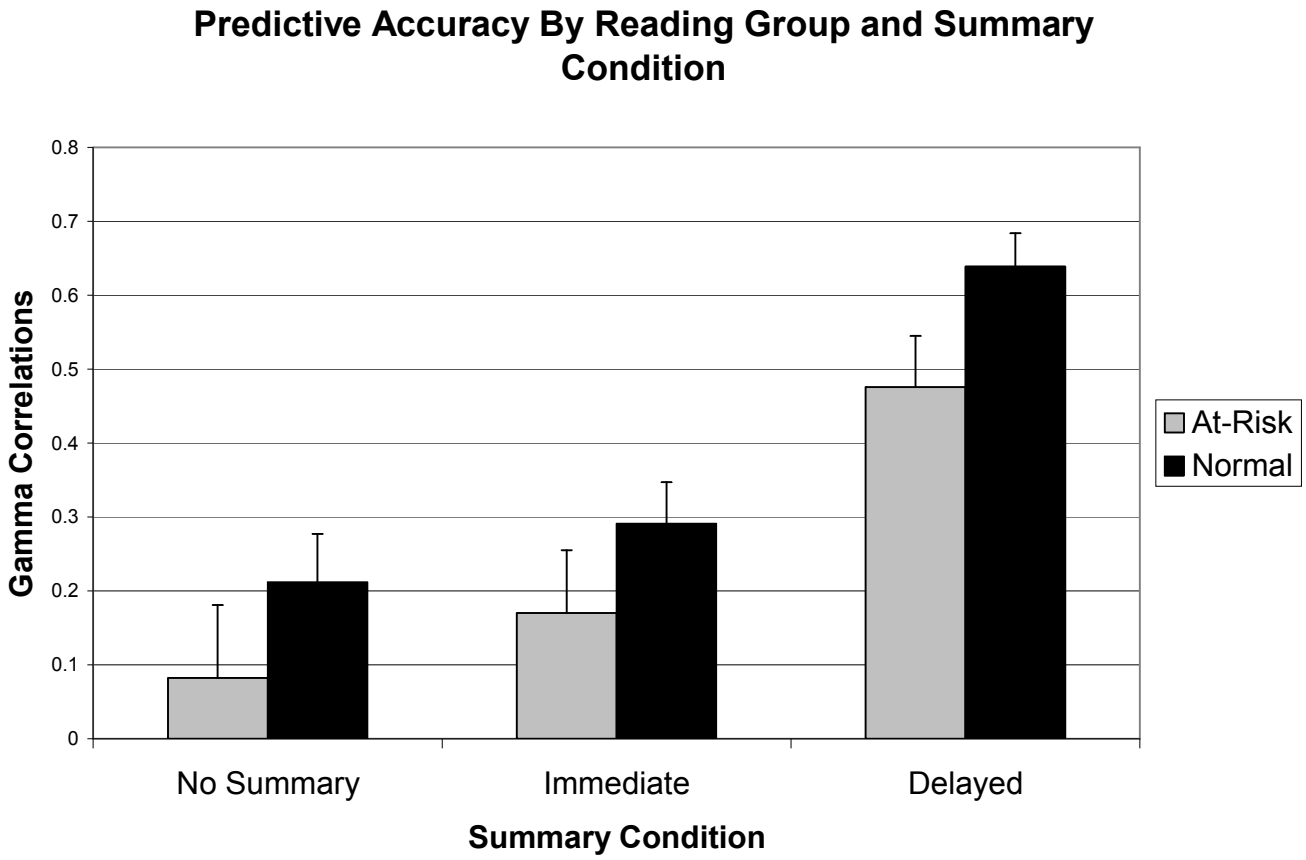
Figure 3. Metacomprehension accuracy by cue use profile and summary condition.

Figure 4. Metacomprehension accuracy by best cue reported and summary condition.

Figure 5. Metacomprehension accuracy for at-risk readers by condition in Experiment 2.

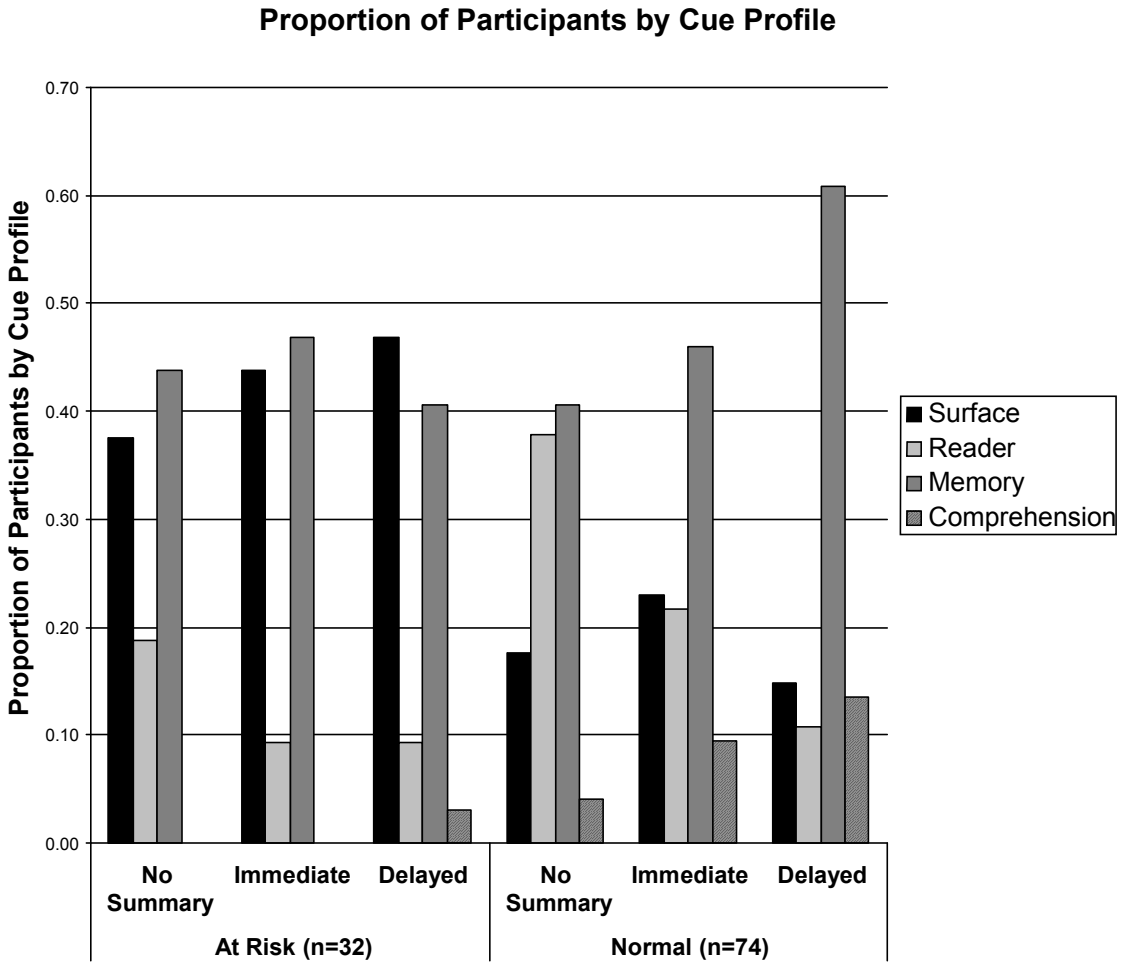
This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Figure 1



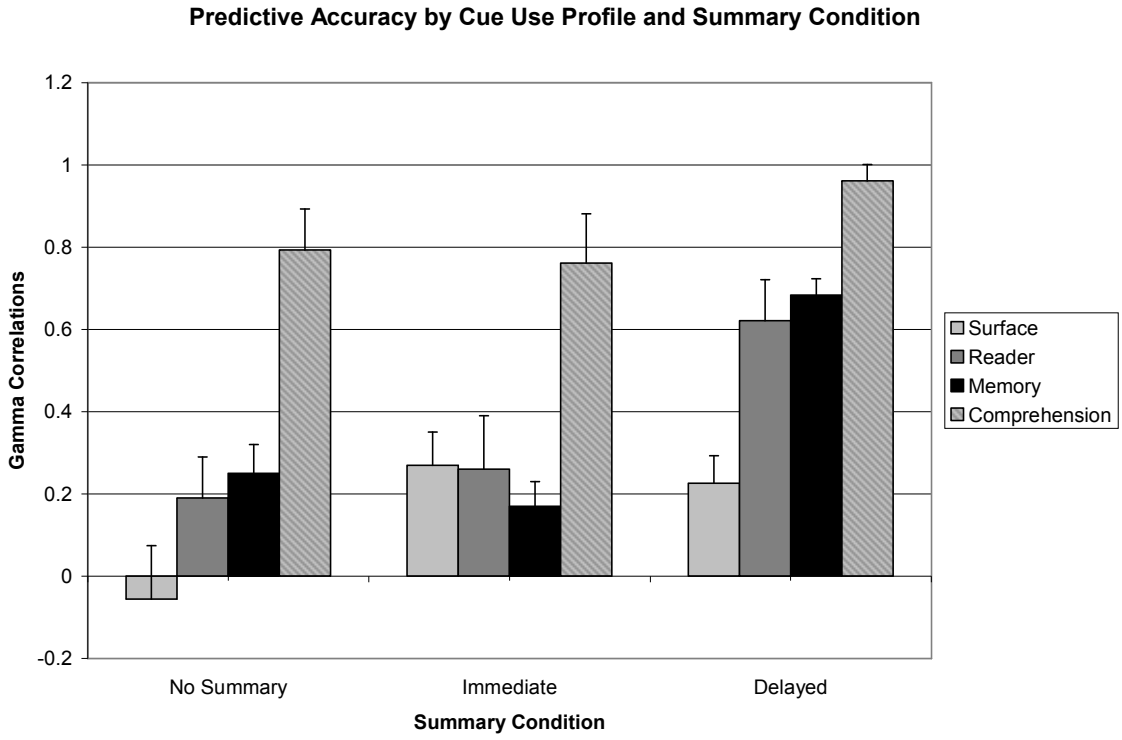
This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Figure 2



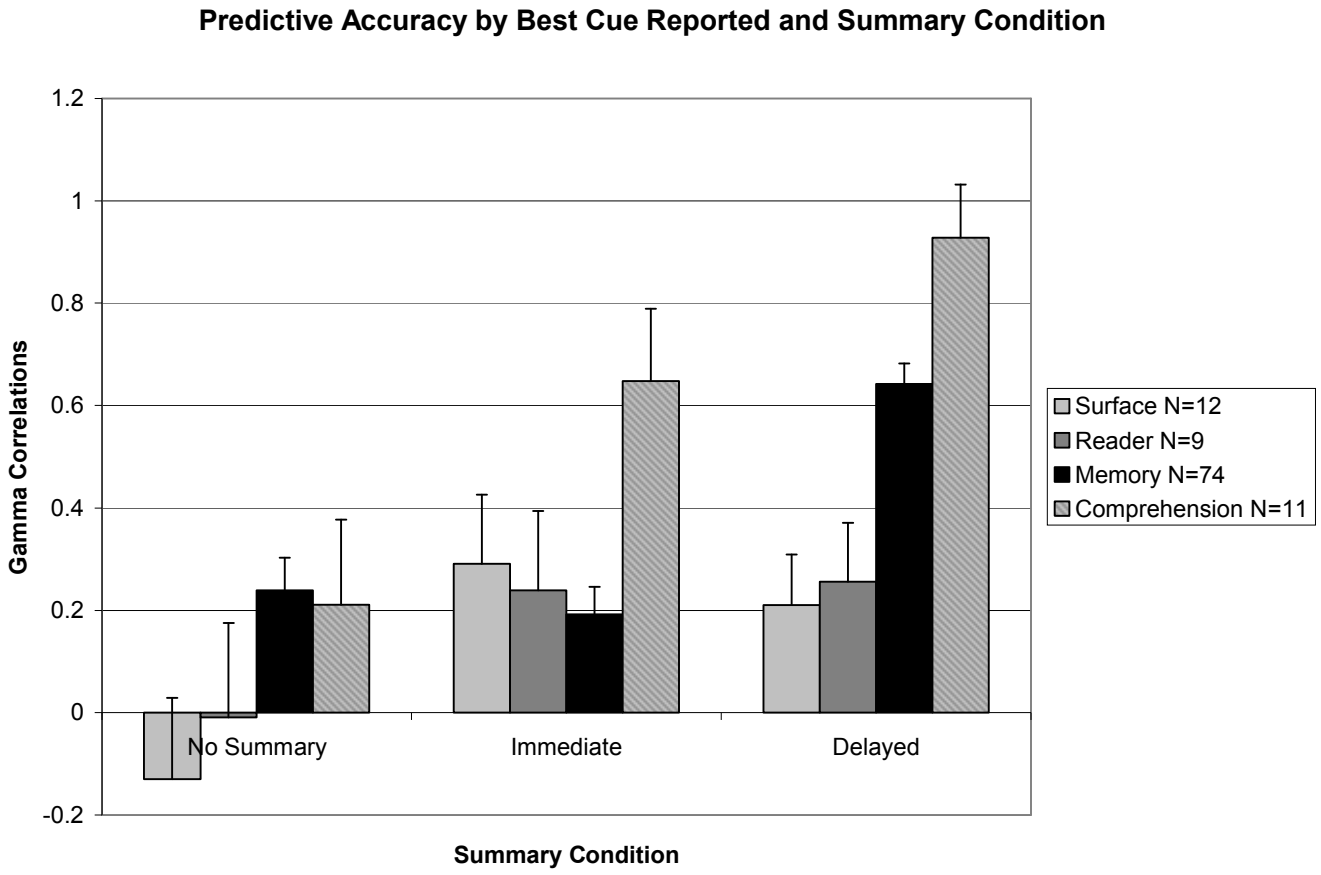
This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Figure 3



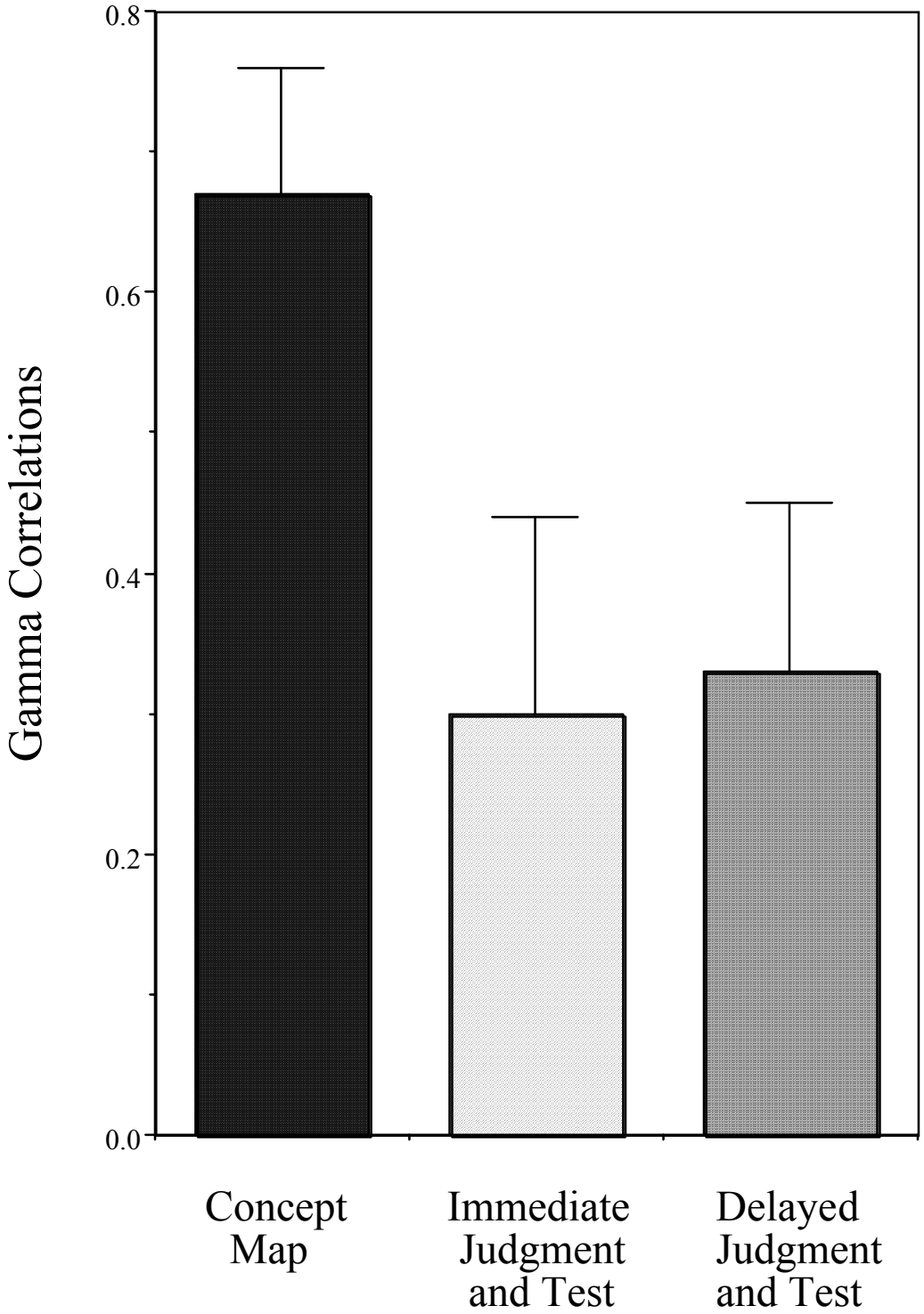
This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Figure 4



This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Figure 5



This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Appendix

Experiment 1: Sample Text

One of the major processes that takes place in schools, of course, is that students learn. When they graduate from high school, many can use a computer, write essays with three-part theses, and differentiate equations. In addition to learning specific skills, they also undergo a process of cognitive development wherein their mental skills grow and expand. They learn to think critically, to weigh evidence, to develop independent judgment. The extent to which this development takes place is related to both school and home environments.

An impressive set of studies demonstrates that cognitive development during the school years is enhanced by complex and demanding work without close supervision and by high teacher expectations. Teachers and curricula that furnish this setting produce students who have greater intellectual flexibility and higher achievement test scores. They are also more open to new ideas, less authoritarian, and less prone to blind conformity.

Unfortunately, the availability of these ideal learning conditions varies by students' social class. Studies show that teachers are most demanding when they are of the same social class as their students. The greater the difference between their own social class and that of their pupils, the more rigidly they structure their classrooms and the fewer demands they place on their students. Students learn less when they come from a social class lower than that of their teacher. The social class gap tends to be largest when youngsters are the most disadvantaged, and this process helps to keep them disadvantaged.

Experiment 1: Sample Test Items

The author probably believes that

A. teachers often come from a lower social class than their students.
B. teachers of the disadvantaged should be familiar with the social class of their students.
C. the social class of teachers and students is of little importance.
D. teachers should be hired who are from a higher social class than their students.

The author seems biased in favor of

A. teachers who are less demanding in working with students.
B. discouraging intellectual flexibility in schools.
C. encouraging students to think critically.
D. giving students less homework.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Experiment 2: Sample Text

The Industrial Revolution refers to the social and economic changes that occurred when machines and factories, rather than human labor, became the dominant mode for the production of goods. Industrialization occurred in the United States during the early and mid-1800s and represents one of the most profound influences on the family.

Before industrialization, families functioned as an economic unit that produced goods and services for its own consumption. Parents and children worked together in or near the home to meet the survival needs of the family. As the United States became industrialized, more men and women left the home to sell their labor for wages. The family was no longer a self-sufficient unit that determined its work hours. Rather, employers determined where and when family members would work. Whereas children in preindustrialized America worked on farms and contributed to the economic survival of the family, children in industrialized America became economic liabilities rather than assets. Child labor laws and mandatory education removed children from the labor force and lengthened their dependence on parental support. Eventually, both parents had to work away from the home to support their children. The dual-income family had begun.

During the Industrial Revolution, urbanization occurred as cities were built around factories and families moved to the city to work in the factories. Living space in cities was crowded and expensive, which contributed to a decline in the birthrate and to smaller families.

The development of transportation systems during the Industrial Revolution made it possible for family members to travel to work sites away from the home and to move away from extended kin. With increased mobility, many extended families became separated into smaller nuclear family units consisting of parents and their children. As a result of parents' leaving the home to earn wages and the absence of extended kin in or near the family household, children had less adult supervision and moral guidance. Unsupervised children roamed the streets, increasing the potential for crime and delinquency.

Experiment 2: Sample Test Items

What is the relationship between these sentences from the last paragraph? "With increased mobility, many extended families became separated...." and "As a result of parents' leaving the home...."

A. cause and effect
B. generalization and example
C. statement and clarification
D. summary

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

From this passage, you can conclude that

A. many of the problems with American families came about since the Industrial Revolution.
B. children who lived on farms were less mature and independent than those reared in the cities.
C. the Industrial Revolution led to stronger and larger American families.
D. improved means of transportation encouraged mothers to stay home with young children.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Dr. Rapp,

Thank you for the wonderful feedback regarding our paper (Ms. No. DP-D-07-00066; Poor Metacomprehension Accuracy as a Result of Inappropriate Cue Use Discourse Processes). We have addressed all of the concerns raised by the reviewers and yourself. To assist you in tracking our changes, we provided your cover letter and the reviews and included our responses within the letter.

Ref.: Ms. No. DP-D-07-00066
Poor Metacomprehension Accuracy as a Result of Inappropriate Cue Use
Discourse Processes

Dear Dr. Thiede,

Thank you for submitting your manuscript "Poor Metacomprehension Accuracy as a Result of Inappropriate Cue Use" to Discourse Processes. We have filed your manuscript under the identifier DP-D-07-00066.

I have now received reviews of the paper from three experts who reflect Discourse Processes' multidisciplinary readership and, I believe, are uniquely qualified to evaluate the work. All of the reviewers find the topic very interesting and potentially relevant for our readership. While they each, in their own way, suggest that the work needs revision before it would be publishable, they all agree that a revised form of the paper would make a nice addition to the journal. Upon my own reading, I agree with the reviewers. Thus, I am happy to accept the paper for publication in Discourse Processes, pending revisions that address the reviewers' concerns.

The reviewers' comments are appended below. I would like you to attempt to address them directly in your manuscript or in your cover letter, as appropriate. All of their comments are useful and informative, and in many cases the comments converge on similar themes. Let me highlight some of those themes here:

(1) The reviewers raised questions about the connections between Experiments 1 and 2. Their concerns are related to differential findings across the experiments (e.g., Reviewer 2), suggestions as to the inclusion of appropriate control comparisons within Experiment 2 (e.g., Reviewer 1), and explicit claims derived from cross-experimental discussions without adequate empirical/statistical support. Please determine the best way to deal with each of these issues to remediate the concerns. This might involve tempering your claims, running an additional comparison group in your study, appealing to previous work, etc. - I leave it to you as to how best to deal with these important issues.

Regarding the differential findings across experiments, we believe the findings across experiments are consistent with the notion that metacomprehension accuracy increases as participants base metacognitive judgments on cues that are related to the situation model of the text. However, we eliminated all discussion of levels of

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

metacomprehension accuracy across experiments. Moreover, when we did comment on the levels of accuracy, we clarified that the comparison was being made to the literature rather than between our experiments.

Regarding the inclusion of a control group in Experiment 2 (i.e., typical readers), this experiment was conducted in a remedial reading classroom. Given that there are not remedial reading courses for typical readers, it's not clear what group would be an appropriate control group. However, we took care not to make comparisons between at-risk and typical readers in Experiment 2 (as typical readers were not represented).

(2) Each of the reviewers called attention to the nature of the self-report methodology you implemented in your experiments, as a means of identifying participants' cue usage. Some discussion of the limits of self-reports in general is warranted, and you will likely want to describe how those limits necessitate qualifications of your claims.

We discussed the limitation of self-reports on page 14 and discuss how we tried to reduce the limitations of self-reports in our conclusions. Specifically, we tried to reduce the limitations of the self-report data by focusing on how self-report data is related to metacomprehension accuracy(which was suggested by Ericsson and Simon, 1980)—rather than on the solely on the self-report data, which could be less accurate.

Please note that although the frequency of self-reported cue use might reflect a bias on the part of participants toward over stating their cue use, overstating cue use should have been fairly consistent across the three within-subject conditions. Moreover, it should have at worst created additional random variability that would deflate the correlation between self-reported cue use and metacomprehension accuracy.

(3) Reviewer 1 raises several important concerns with respect to your categorizations of participants' responses and the resulting comprehension profiles. Your discussion of the results should address these concerns, perhaps with reflection on the design of your coding schemes. Also, please evaluate your usage of proportion and frequency data in presenting your results/figures and making your claims.

The design and justification for the coding schemes has been revised and elaborated on pages 17-19. The new figure 2 presents the cue use categorization as a proportion of each reading group, but importantly the sample sizes are listed as well, which was an important aspect of presenting the data as frequencies.

(4) Reviewer 3 brings up several ways in which the Results sections might be tightened up to present the data in a more digestible form.
I similarly made a note during my reading in several sections (most notably, with respect to the rather lengthy "cue quality" discussion on pages 19-21). I'm hoping you will be able to make adjustments that enhance the clarity of your data presentations without unnecessarily increasing the length of the manuscript.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

We provided more foregrounding of our analyses to indicate what is replication and what is new. We also provide a better overview of why alternative analyses are needed.

(5) While each reviewer expressed enthusiasm about the inclusion of at-risk students in your project design, they also felt that the introduction and theoretical justification for their inclusion in the project was relatively thin. I would encourage you to enhance the introduction of this issue beyond, based on my reading, the single paragraph on page 7.

We added a paragraph highlighting our reasons for including at-risk readers, see the bottom of page 7 and top of page 8. on page 7.

Minor comments from my own notes:
-You might provide a brief explanation of gamma correlations for the readership, perhaps in a footnote.

We added a footnote describing gamma on page 3.

-The categorical coding completed by research assistants should, generally speaking, be reported as kappa for your inter-rater reliability analysis, as a pure percentage of agreement does not take into account the possibility of agreement by chance. It is also important to provide an indication of what the coders did with non-agreement cases.

We report kappas on pages 14 and 34.

-On page 21, I believe Figure 3 should be Figure 4; on page 30, Figure 4 should be Figure 5.

We made these changes.

-The claim on the top of page 6 with respect to the decay rates of surface versus situational representations, while informative, is not entirely the same as suggesting that those representations are necessarily utilized in an analogous temporal pattern. An additional statement or two concerning usage rather than representation, with a citation or two, would be helpful for making this connection.

We do not understand your point. Our claim is if surface memory for text no longer exists after a delay, it cannot be used as a basis for judgments. The citations for this point are already cited: Kintsch, Welsh, Schmalhofer and Zimny (1990; see also Fletcher & Chrysler, 1990; Schmalhofer & Glavanov, 1986).

We added the following to clarify our point:This interpretation is based on the work of Kintsch, Welsh, Schmalhofer and Zimny (1990; see also Fletcher & Chrysler, 1990; Schmalhofer & Glavanov, 1986) which has shown that access to surface information decays rapidly, whereas access to the situation model is more robust over time. This would mean that surface memory for text would be less accessible after a delay, and

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

thus less likely to mislead readers as a basis for their comprehension judgments.-On the bottom of page 4 to the top of page 5, the argument is that comprehension tests tap situation models. I think some notion of the TYPES of VALID comprehension tests that tap situation models would be valuable, so as not to confuse readers that all tests unerringly and uniformly do so.

Our point here is IF comprehension tests tap the situation model, then readers need to use situation model based cues. This has been clarified.

ACTION: I am accepting this paper for publication pending revision.

When you send in your revision, I may send it back to some of the current reviewers for their re-evaluation of how it has addressed their concerns. In submitting your revision, please complete the following steps:

1. Complete your revised manuscript as indicated (reviewers' comments are appended below).
2. Go to http://dp.edmgr.com/ and log in as an Author. When you reach the main menu, you will find your submission record by clicking on Submissions Needing Revision.
3. Click Submit Revision and begin following the same steps you did in your original submission.
4. In submitting your revised file(s), please attach your revised manuscript (and any revised figures or tables). Also, please provide a cover letter file detailing how you have addressed the reviewers' concerns.

Please ensure that the entire manuscript conforms to the style guidelines in the Publication Manual of the American Psychological Association.

If you have any questions along the way, please feel free to contact me at rapp@northwestern.edu.

Thank you for considering Discourse Processes as an outlet for your work.

Sincerely,

David N. Rapp, PhD
Associate Editor
Discourse Processes

Comments from Reviewers:

Reviewer #1: Review for Discourse Processes - Manuscript DP-D-07-00066
Title: Poor Metacomprehension Accuracy as a Result of Inappropriate Cue Use

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Summary: Two experiments were conducted to test the "situation model approach to metacomprehension" that was proposed by the authors in a previous article. The first experiment investigated the cues that typical and "at-risk" readers report using when judging their comprehension and attempted to establish that, for both groups, cues pertaining to a reader's situation model lead to the highest levels of metacomprehension accuracy. The results showed that self-reported use of situation model cues (i.e., comprehension- and memory-based cues) was associated with increased metacomprehension accuracy, but only when reading was followed by a summarization manipulation. The results also showed that "at-risk" readers were less likely than typical readers to report using situation model cues and more likely to report using surface cues, which were associated with low levels of metacomprehension accuracy. Thus, the purpose of the second experiment was to investigate a reading intervention that would potentially decrease "at-risk" readers' reliance on surface cues and increase their use of situation model cues. More specifically, participants were asked to construct a concept map for each text before rating how well they understood it. The results showed that constructing a concept map (much like writing a summary) led to increases in metacomprehension accuracy.

Review: This paper serves two important, but somewhat disparate purposes. First, it examines people's beliefs about the kinds of strategies they use to make metacomprehension judgments and, second, it demonstrates the effectiveness of concept mapping as a means of increasing metacomprehension accuracy. The first purpose is of particular interest because it adds to the growing body of literature concerned with documenting and evaluating (in terms of accuracy) the heuristics that underlie people's judgments of comprehension. For the most part, this research has identified heuristics by manipulating the cues (e.g., domain familiarity, ease of processing, etc.) on which people base their judgments. Experiment 1 is the first study that I know of to undertake the arduous but worthwhile task of asking people what heuristics they use when judging their comprehension. Although I disagree with some of the conclusions drawn from the results, I think the experiment provides convergent validity for many of the previous findings in the literature. It also begins to answer the question of whether people have introspective access to the kinds of heuristics they use. As for the second purpose of the paper: identifying yet another intervention capable of increasing metacomprehension accuracy is important, but perhaps of less interest to reading comprehension researchers than to educators. In fact, it was hard to see how Experiment 2 really fit with the purposes of Experiment 1. For this reason (and others), I think the paper needs to be significantly revised. Below, I discuss some of the major issues that I think should be addressed in such a revision:

1) If the purpose of Experiment 2 was to demonstrate that a concept mapping intervention can "provide a context for at-risk readers that may give them direct access to valid cues for judgment" (p. 23), then why were participants not asked about what cues they used to make metacomprehension judgments, as in Experiment 1? It is true that concept mapping increased the metacomprehension accuracy of at-risk readers, but we do not know if this because they attended more to "comprehension-based" cues or because it made the cues they were already using more diagnostic of actual comprehension. Based

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

on Experiment 1, it seems that the latter is more likely to have been the case. That is, although there were no differences in the distribution of cue use across summary conditions for at-risk readers (p. 16), their metacomprehension accuracy was higher in the delayed-summary condition than it was in the immediate- and no-summary conditions.

We did not think to add the self-report to the design of Experiment 2, as we did not know the results of Experiment 1 when it was being designed/run. Because this was an in-class intervention, we wanted to keep the study as simple and short as possible so as to not overload the at-risk readers.

In the introduction and discussion of Experiment 2, we acknowledged the alternative explanation proposed by Reviewer 1 (i.e., accuracy changed because the validity of cues changed from one situation to another).

We do, however, think we have some evidence that concept maps gave students valid cues. In particular, as reported on page 34, the number of connections contained in concept maps was predictive of performance on tests of comprehension AND the number of connections was also related to metacomprehension judgments.

Which leads to my next point: we already knew from Experiment 1 that delayed summarization was an effective intervention for increasing the metacomprehension accuracy of at-risk readers, so, for the purposes of this paper, why was it necessary to demonstrate that concept mapping also served as an effective intervention? Perhaps, as the paper suggests, it is because the lack of a Reading Group ×Summary Condition interaction for metacomprehension accuracy indicated that "the delayed summarization instruction was not a strong enough intervention to equate the accuracy of the two reading groups" (p. 15). But if this is the reason, then why did Experiment 2 not include either a sample of typical readers or a delayed summary condition? It is impossible to tell from the results of Experiment 2 whether concept mapping was strong enough to equate the accuracy of at-risk and typical readers or whether concept mapping was a stronger intervention than delayed summarization. Although the fact that concept mapping increases metacomprehension accuracy among at-risk readers is an important finding, I do not think it contributes to the general goals of the paper. Thus, it may be worth focusing just on Experiment 1.

We changed the introduction to Experiment 2 to include the explanation that metacomprehension improved because the validity of cues changed from one situation to another. We also clarified that the goal of Experiment 2 was to change the cues used by at-risk readers.

2) The paper claims that asking readers to report the cues they used to judge their comprehension provides the first "direct investigation" of whether interventions such as delayed summarization "shift readers from monitoring poor cues to better cues for predicting their own comprehension" (p. 6). However, the validity of retrospective self-reports as direct measures of psychological processes (especially the processes that

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

underlie our judgments) has long been called into question (e.g., Ericsson & Simon, 1980; Nisbett & Wilson, 1977). Although some of the participants in Experiment 1 may have accurately reported the cues they used when making comprehension judgments, I doubt this was the case for most participants - especially since metacognitive judgments are often based on implicit processes (e.g., heuristics such as cue familiarity that lead to feelings of knowing; Reder & Schunn, 1996). A more direct test of whether a particular cue was used more frequently in the delayed-summary condition than in the no-summary condition would have been to manipulate the salience, availability, or diagnosticity of that cue (e.g., Rawson & Dunlosky, 2002).

Past investigations have used manipulations that were assumed to affect the availability of cues, without actually measuring cue use. Cue use was inferred from differences between conditions. What we meant by direct investigation is that we attempted to get an actual measure of cue use via self report. We admit self-report data can be problematic, and note this in the manuscript on page 14, but it has been shown to be useful in some contexts, such as asking students about their memory strategies. Moreover, in our study, we have evidence that the self-report measures were valid, because reports varied across condition (for typical readers), and predicted accuracy in a systematic and predictable way.

3) The coding system for categorizing participants into one of four cue-use profiles does not have a clear rationale (p. 15). Why were participants who reported using any cues related to the qualities of the text itself classified as fitting the surface profile, regardless of any other cues they reported? This seems to be contrary to the logic that was later followed for the best-cue analysis (p. 19). That is, why were participants not categorized "as a function of the highest quality cue" they reported using? By this logic, participants who reported using a text-based cue and a memory-based cue should have been classified as fitting a memory profile.

This logic has been better explicated as noted in the response to the action letter.

4) Participants in Experiment 1 "who reported relying on their ability to understand or explain the text were classified as using comprehension-based cues" (p. 16). However, this definition of comprehension-based cues seems to beg the question of what cues participants used to judge whether their understanding of the text was sufficient. It may be that participants who were classified as fitting a comprehension profile relied on the same cues as participants who fit a surface or reader profile and that the only real difference is in the way participants in each category described their use of these cues. The example of a comprehension-based cue presented in the paper was "I gave [my comprehension] a high number if I thought I could explain the meaning of the story to another person" (p. 12). But, perhaps this participant believed she could explain the meaning of the text because it was not particularly difficult for her to read (i.e., she experienced a feeling of fluency while reading). This would means that she actually

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

relied on a reader or surface cue.

We admit that there are issues with self-report data, and that the way people express their ideas introduces random error, but we had to take the comments at face value for coding purposes. To interpret them beyond what was stated would be problematic. Again, there is evidence that the reports were valid in our data set as they were predictive of accuracy.

5) Related to the previous point, is a relatively minor concern about the way reading groups were described as following either a "heuristic" or "metacognitive" approach (p. 20). I do not agree with the claim that surface- and reader-based cues are simply heuristic approaches that "do not actually require 'meta' awareness." When a person has a sense that the content of a particular text feels familiar or that the text was relatively difficult to read, how is this any less "meta" than when a person senses that she will be able to remember "the ideas from the article"? Both sets of cues are the output of judgment heuristics and both require the individual to have a sense of their own understanding (as an entity or state that can be judged). Thus, what makes all of these cues metacognitive is that they allow people to make inferences about their own cognitive states.

The distinction we were trying to make was between externally available information that is used for monitoring vs. monitoring based on privileged internal representations. However, we agree that both can be 'meta" and have re-written this section on page 23 to remove the argument of which cues are and are not meta.

My remaining concerns are less general and thus will be discussed in order of appearance:
a) In the second line of the abstract, I think it would be more appropriate to say that the studies identify "interventions," not "learning contexts."

We made this change.

b) Towards the top of p. 3, the term "monitoring accuracy" needs to be explained (as on p. 14).

We did this in footnote 1.

c) The last paragraph of p. 6 suggest that the paper will provide evidence for the claim that interventions such as delayed summarization "shift readers from monitoring poor cues to better cues for predicting their own comprehension." However, the discussion of Experiment 1 suggests that the intervention is effective because of a shift in the validity of existing cues (particularly memory-based cues), not a shift from one type of cue to another.

We revised this on page 6 to reflect that improvements in accuracy could be due to changes in the validity of cues.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

d) The explanation of why the experiments focused on at-risk readers (top of p. 7) seems a bit ad-hoc. What does using a sample of at-risk readers have to do with the purposes of the paper described in rest of the introduction (e.g., testing the situation model approach)?

We clarified our motivation for using at-risk readers in the Introduction (pages 7-8).

e) The opening sentences of the intro to Experiment 1 (p. 7) do not describe what seem to have been the primary purposes of the experiment (i.e., testing the situation model approach).

We added this as a primary goal of Experiment 1.

f) In the materials section on p. 9, it should say that there were 5 text in each of the three sets.

We added this.

g) It seems a bit problematic that the 10 multiple-choice test questions did not include any memory-related items (p. 9). It would have been interesting to see whether (compared to comprehension-based cues) memory- or reader-based cues were associated with higher accuracy for memory questions (see Thomas & McDaniel, 2007a, 2007b).

We intentionally did not use memory items because we think it confuses the reader and gives them incorrect expectations, as we explicated in another paper (Wiley et al. 2005). We are currently working on another paper that directly tests the effects of combining versus isolating test items at different levels of representation, but this issue is beyond the scope of the current paper.

h) Perhaps, on p. 10, there should be a discussion of why the typical JOL scale (predictions of future performance, 1-100) was not used. This could also be a footnote.

We explain on page 11 that we used the same prompts as Glenberg and Epstein (1985) –the original study in metacomprehension.

i) In terms of the coding scheme on p. 12, where would ease of processing or fluency fit in?

This phrase has been added to the coding justification.

j) Why did typical readers perform worse in the immediate summary condition than in the other two conditions (middle of p. 14)? It may be worth offering a possible explanation.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

This was a very small effect, which is not relevant to the main goals of the paper. We don't believe this effect can be easily explained and are reluctant to dedicate space to speculate as to the cause of the effect.

k) Why did typical readers report using memory-based cues more often in the summary conditions than in the no-summary condition (bottom of p. 16)?

We provide an explanation for this effect in our discussion of Experiment 1 when we say"Thus, when readers base their cues on their ability to remember a text, which become apparent during a summarization task, their judgments will be more predictive of comprehension test performance as long as some time passes after reading but before attempting to summarize. The present data provide support for this account."

l) In each experiment (p. 8, p. 27), what incentives did participants have for completing the study? Were they paid, did they receive extra credit, was it a course requirement?

We added that participation was part of the course requirements (see pages 9 and 31).

Reviewer #2: This paper reports two experiments focusing on cues that enhance the accuracy of metacomprehension. In Experiment 1, at-risk and typical readers indicated what cues they had used in making metacomprehension judgments. At-risk and typical readers differed in how frequently they made use of memory- and comprehension- related cues. They made more use of such cues in a delayed summary condition, and metacomprehension was more accurate in that condition. Interestingly, the two types of readers were equal in metacomprehension accuracy when they made use of the same higher-level cues. In Experiment 2, at-risk readers were trained to use concept maps during reading. Their metacomprehension judgments in the concept map condition were very accurate. Overall, I believe that these findings are novel and important. Although I would recommend some revisions, the experiments will make a nice contribution to the literature.

I worried about the use of self-report for determining which cues were used. Can participants actually report what they were doing accurately? There isn't anything the authors can do about this, but some discussion of why self-reports are likely to be valid in this situation might be helpful.

As noted above, a discussion of the limitations of verbal reports and our reasons for believing they are valid in this situation is now included

The fact that delayed summaries helped both typical and at-risk readers equally is interesting. It is also interesting that there was a main effect of reading ability. Such individual differences have not been easy to find in the metacomprehension literature.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

Figure 2 needs to be re-made with percentages. Because there were very different numbers of participants in the typical and at-risk groups, it is difficult to compare how often each group used the various cues. The differences are not as large as one might expect from the figure.

The figure has been revised.

For the analysis on p. 17, how could there be four levels of cue use in both groups for the No Summary condition? In Figure 2, it appears that no at-risk participants used the comprehension cues (same for the immediate-summary condition).

There is an empty cell in this analysis. The main effect for cue use profile was computed by summing across the levels of reading group (eliminating the empty cell in the computation). However, the interaction is not reported due to this missing cell.

The results of Experiment 2 seem to be somewhat inconsistent with those of Experiment 1, presumably because different materials were used? In the no summary conditions, the at-risk readers produced gammas of about .3 in Experiment 2, but .09 in Experiment 1. Indeed, the at-risk readers in Experiment 2 produced higher gammas than did the typical readers in the no-summary condition of Experiment 1. The authors should address this- are these differences simply due to differences in materials?

That is one possibility. Experiment 2 was also run in an intact classroom setting and administered by the teacher, as opposed to Experiment 1 which was run as a subject pool experiment. Because we are trying to eliminate between-experiment comparisons as directed by the action editor, we do not address it in the paper.

It is unfortunate that a typical reader control group was not tested in Experiment 2. Granted the gamma in the concept map condition was very high, but all of the gammas in Experiment 2 were high relative to Experiment 1, so comparing across experiments is risky.

We agree, so we did not make the comparison across experiments.

Some minor comments and suggestions are below:
p. 5- the Dunlosky & Lipko reference is not in the reference list. The paper is already published and not in press, I believe.

We added this reference.

On p. 18, the results of the analysis should be clarified. Monitoring accuracy was significantly worse for readers who fit a surface cue profile than what? Accuracy was

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

better for readers who fit a comprehension-based profile than what? Is it these two groups that are being compared here?

In each ANOVA only the main effect for cue use is significant. Therefore, the follow up tests that are reported are comparing each cue use profile to each other profile group. Significant differences between sets of groups as well as groups that do not differ from each other are noted.

p. 22- Summary from Cue Analysis- the comprehension based cues were the best predictors for delayed comprehension tests.

Corrected

p. 23- middle- hypothesis "led" not "lead"

We correct this typo.

p. 24- short paragraph- did not affected.

We corrected this typo.

I assume that the participants in Experiment 2 had not participated in Experiment 1 (if they did, maybe that's why they were so much better in Experiment 2).

They did not and we added this to the Participant section of the Method in Experiment 2.

The tables should be placed in front of the figures in the manuscript, and the Appendix goes earlier also.

We did this.

Reviewer #3: Poor Metacomprehension Accuracy as a Result of Inappropriate Cue Use

The study reported in this study continues the productive line of research of Thiede and his associates on metacomprehension. Previous research on metacomprehension yielded the following pattern: First, accurate metacomprehension monitoring improves the effectiveness of study regulation, such as the choice of which items to re-read. Second, in turn, effective regulation of study improves overall reading comprehension. Finally, however, monitoring accuracy is by and large quite poor. Thiede and his associates devised several clever manipulations that improve reading monitoring, leading ultimately to improved test performance. The present study is based on the idea that the effectiveness of these manipulations derives from the fact that they induce readers to base their judgments of comprehension on cues that are related to what Kintsch calls situation-model level rather than on those that tap surface level that readers us for comprehension judgments, co because surface information decays rapidly over time.

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

The manuscript certainly deserves publication. The questions are framed in terms of issues and ideas that are at the focus of the text-comprehension literature, and can be of interest to readers of Discourse Processes, The inclusion of a comparison between a group of typical college readers and a group of students who are required to attend remedial reading classes is very important and yielded interesting findings about the monitoring deficits of the latter group. Also the collection of self-report data on the bases judgments of comprehension allows the authors to obtain more direct evidence for some of the ideas that have received only indirect support in previous studies. Finally, although Experiment 2 is demonstrative in nature, its results are promising and I am sure that they will lead to important applications.

The article as a whole is well written. There are a couple of recommendations that I would make, however. First, because some of the effects observed replicate those of the previous studies (e.g., the beginning of the Results section of Experiment 1), I would suggest writing the results sections by first focusing on the trends that replicate previous findings (also mentioning again these previous findings) and then adding the new observations. This will make it easier for readers who are not familiar with the previous literature to obtain a clearer picture. For such readers, I would also recommend giving a few concrete facts at the very beginning of the article regarding the deplorable, low metacomprehension accuracy that has been observed in previous studies. I would also suggest adding a reference to Dunlosky and Lipko's (2007) recent review of metacomprehension research in Current Directions in Psychological Science.

We added this reference to the discussion on the bottom of Page 3 as well as a mention that metacomprehension accuracy is typically dismal.

Second, self-report data are clearly important in providing some insight into the bases of metacomprehension judgments. However, people are not always aware of the cues that they use. I suggest that the authors mention this reservation.

See above

Also, it is not appropriate to speak of "the effects of cue use" on metacomprehension accuracy, because the results are correlational.

We removed this language from the results of Experiment 1.

Third, the analyses of the self-report data (Experiment 1) are very difficult to follow, although some of the trends observed are clearly interesting and important. I do not have a clear suggestion how to improve presentation. One possibility, perhaps, is to report just one analysis and then examine how alternative analyses agree with it. Another possibility, which is not ideal, is to see whether analyzing the results "backwards" can provide a better picture: Divide participants into high and low in metacomprehension accuracy (by

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927

condition) and then compare the cues that they report. I should say that the Discussion section of Experiment 1 is very well written, and helps provide an overview of the findings.

We provided an overview of these analyses, which helped clarify the purpose of each analysis.

All in all, this is an important manuscript that deserves publication.

Again, we want to thank you and the reviewers for the excellent feedback. We hope we have addressed all the concerns adequately and hope the manuscript is acceptable.

Thank you,

Keith Thiede, Jennifer Wiley, Thomas Griffin, and Mary Anderson

This is an electronic version of an article published in Discourse Processes, 47(4). Discourse Processes is available online at: http://www.informaworld.com/smpp.
DOI: 10.1080/01638530902959927