Metacognition Learning (2012) 7:113–131
DOI 10.1007/s11409-012-9087-y

The effect of metacomprehension judgment task on comprehension monitoring and metacognitive accuracy

Yasuhiro Ozuru · Christopher A. Kurby · Danielle S. McNamara

Received: 4 March 2011 / Accepted: 25 April 2012 /
Published online: 10 May 2012

Abstract The authors investigated differences in the processes underlying two types of metacomprehension judgments: judgments of difficulty and predictions of performance (JOD vs. POP). An experiment was conducted to assess whether these two types of judgments aligned with different types of processing cues, and whether their accuracy correlated with different factors such as sensitivity to processing ease and reading ability. Participants (n=72) read an extended text about brain structure and after each sentence made either a JOD or POP. Results suggested that JODs and POPs were made based on different sets of cues because different factors correlated with the accuracy of metacomprehension judgments. JOD accuracy correlated with sensitivity to processing ease and POP accuracy most strongly correlated with reading ability. Engaging in different metacomprehension judgments during reading may alter the information sources to which a reader attends and which factors influence metacognitive accuracy.

Keywords Metacomprehension · Science text · Comprehension judgment

Recent research on metacomprehension has indicated that comprehension monitoring performance may depend on specific reading situations and study strategies, such as making judgments of learning, judgments of difficulty, reading to solve specific questions, and delayed summaries (e.g., Thiede and Anderson 2003; Thomas and McDaniel 2007; Thiede et al. 2009). This study examined how two commonly used judgment strategies in metacomprehension differ in terms of the information source they draw from (i.e., cue) and their resulting metacomprehension accuracy. These metacomprehension strategies are:

Y. Ozuru (✉)
University of Alaska Anchorage, Anchorage, USA
e-mail: afyo1@uaa.alaska.edu

C. A. Kurby
Grand Valley State University, Allendale, USA

D. S. McNamara
Arizona State University, Tempe, AZ 85287, USA

judgments of difficulty (JOD) in which readers monitor comprehension by judging how difficult it feels to understand a text as they read, and *predictions of performance (POP)* in which readers monitor their comprehension by thinking about how well they would perform on future test question(s) based on the text they are reading.

There has been a good deal of research on how well people consider they have learned new material (judgments of learning: Dunlosky and Bjork 2008). Recent work has extended this line of investigation to the study of reading comprehension, focusing on task and individual differences (see Thiede et al. 2009, for a review). But, how readers accomplish this monitoring process (e.g., what information is used) from moment-to-moment during text comprehension is still not well understood. Further, the way readers monitor their comprehension from moment-to-moment likely differs depending on the type of judgment task or situation. For example, readers may rely on a separate set of cues when judging their perceived difficulty of comprehension (as in JODs) than when predicting how well information will be remembered on a comprehension test (as in POPs). Further, the different ways of monitoring comprehension may influence the accuracy of a reader's comprehension monitoring. Thus, this study explores how JODs and POPs differ by focusing on: 1) the cues used in metacomprehension; 2) metacomprehension accuracy; and 3) the factors contributing to metacomprehension accuracy during discourse understanding.

Cues used in JOD and POP

JODs and POPs correspond with two of the typical ways in which readers monitor their comprehension. JODs involve an evaluation of comprehension in terms of the processing ease (or difficulty) experienced while reading the text, and this may be the way in which people monitor their understanding when reading for nonspecific purposes such as enjoyment. For example, a reader may abandon a book after reading through the first page and finding the content to be too difficult. Alternatively, a reader may monitor comprehension in a more systematic fashion by thinking about a specific reading goal. For example, using a POP as a monitoring process, readers may evaluate their comprehension with reference to whether they can successfully answer future comprehension questions or write an essay on the topic.

Given that these two reading situations and corresponding metacomprehension processes constitute typical real-life circumstances in which reading takes place, it is important to gain a better understanding of these two types of metacomprehension situations, in particular in relation to the cues used in the judgment. The importance of cues in metacognition has been articulated in the cue-utilization framework proposed by Koriat (1993). This framework states that when the employed cues are diagnostic of future performance, metacognition is expected to be relatively accurate. The cue-utilization framework (Koriat 1993) can be similarly applied to the study of metacognitive monitoring during reading, termed *metacomprehension* (e.g., Dunlosky et al. 2002). According to the cue-utilization framework, metacomprehension accuracy should be affected by the extent to which the cues used to predict/judge comprehension are diagnostic of actual comprehension performance (Thiede et al. 2009). Within this framework, then, it is important to gain an insight into what cues are used in metacomprehension judgments such as JODs and POPs and how the use of those specific cues influences metacomprehension accuracy.

Relatively little research has examined precisely how the cues used in JODs and POPs differ. Further, the findings are somewhat mixed with respect to whether JODs and POPs differentially rely on different sets of cues (Rawson and Dunlosky 2002). For example, Maki

et al. (1990) examined whether judgment ratings of ease of comprehension (similar to JODs) and POPs are influenced by the comprehensibility of text as manipulated by the deletion of letters within paragraphs. Less comprehensible texts should be rated as more difficult. Maki and colleagues found that ease of comprehension judgments were lower for deleted-letter paragraphs than intact ones, whereas POP ratings did not differ between the conditions. Thus, Maki et al. (1990) concluded that ease of comprehension judgments were made based on ease of processing, whereas POPs were not.

Using a similar methodology, Rawson and Dunlosky (2002) reported different results. They examined, in four experiments, whether ease of processing provides an important cue in POPs. These experiments showed that POP ratings were higher when sentence pairs were more causally coherent, when paragraphs had a more coherent structure, or when paragraphs were intact with no deleted letters. More importantly, Rawson and Dunlosky (2002) also showed that the effect of the ease of processing manipulations was similar across ease of comprehension judgments (similar to JODs) and POPs (Experiment 4). These findings led Rawson and Dunlosky (2002) to argue that POPs are based on ease of processing. In sum, the findings from Maki et al. (1990) and Rawson and Dunlosky (2002) conflict regarding whether a reader's perception of processing ease is used as a cue for POPs as well as for ease of comprehension judgments (similar to JODs).

In relation to the Rawson and Dunlosky (2002) findings, reading time has been claimed by some to be a relatively sensitive measure of the moment-by-moment experience of processing ease during online text comprehension (Haberlandt and Graesser 1985). If the Rawson and Dunlosky (2002) results reflect experience of processing ease, then one would expect a correlation between reading time and POP rating. However, they did not observe such a relation. They argued that reading time is likely to serve as a mediator for POPs only for relatively simple material such as single words and paired associates (e.g., Benjamin et al. 1998) as opposed to extended texts. They further argued that POPs may be directly based on processing ease, without mediation of reading time as a cue.

However, the failure to observe a relation between reading time and judgment ratings may be due to several other factors that have been overlooked. First, participants may make POPs using an analytical assessment of the characteristics of the stimulus materials rather than based on one's experience of ease of processing. Second, Rawson and Dunlosky (2002) correlated judgments with reading time that were aggregated over multiple sentences. The aggregated reading time scores may be too coarse to detect possible fine-grained relations between reading time and judgment ratings. Hence, it is not clear from these studies whether readers use different sets of cues in different metacomprehension tasks (i.e., JODs vs. POPs), and in particular, whether the experience of processing ease differentially influences POPs and JODs.

Effect of differential use of cues on metacomprehension accuracy

Based on the literature we have described, we hypothesize that JODs and POPs involve a differential use, or weighting, of cues in metacomprehension judgments. If this assumption is correct, an important question is whether these two types of metacomprehension judgments lead to differences in metacomprehension judgment accuracy.

In this study, we focus on relative metacomprehension accuracy (Maki et al. 2005; Thiede et al. 2009). Relative metacomprehension accuracy represents the extent to which metacomprehension judgments and actual comprehension performance covary (Thiede et al. 2009), and is usually represented by the gamma correlation (Nelson. 1984). Relative accuracy contrasts with absolute accuracy which represents the extent to which a person's

estimate of comprehension of specific information (e.g., a paragraph) corresponds with the actual level of comprehension, and is represented by various measures such as hit rate, overestimation, and underestimation (see Maki et al. 2005, for a discussion on the difference between relative and absolute metacomprehension accuracy).

Prior research indicates that some manipulations affect the relative accuracy of metacomprehension. For example, Thomas and McDaniel (2007) reported that *enriched encoding* improved metacomprehension accuracy. They use this term to refer to the level of encoding resulting from re-generation or elaboration of the text content. Enriched encoding includes self-explanation (Griffin et al. 2008), delayed summary (Anderson and Thiede 2008), diagrams (Cuevas et al. 2002), question generation (Davey and McBride 1986), and summarization (Thiede and Anderson 2003), to name a handful. By contrast, other researchers (e.g., Thiede et al. 2009) refer to these manipulations as relating to the use of "situation model-based cues" with reference to the Construction-Integration model of text comprehension (Kintsch 1998).

These types of manipulations, however, likely cause changes not only in metacomprehension processes but in comprehension processes as well by facilitating deeper text processing. These comprehension processes, in turn, likely render cues that are diagnostic of deeper processing more salient for readers. As such, these types of manipulations increase the likelihood that the cues used for the metacomprehension judgment become more aligned with the cues that are diagnostic of performance on the subsequent comprehension assessment because deeper processing tends to result in better long-term retention of the information.

Alternatively, one can ask whether manipulation of the metacomprehension task alone, as opposed to the manipulation of a comprehension task, as examined in the enriched encoding paradigm (e.g., Thomas and McDaniel 2007), is sufficient to improve relative metacomprehension accuracy as measured by relative accuracy gamma. At least two studies explored this issue, and indicated that relative metacomprehension accuracy was better when participants predicted their test performance as opposed to making simple ease of comprehension judgments (Maki and Serra 1992b; Rawson et al. 2002).

This outcome suggests that processes involved in POPs are more aligned with the processes engaged during test taking than are JODs, perhaps rendering cues involved in POPs more reliable for comprehension monitoring. However, one limitation of the studies conducted by Maki and Serra (1992b) and Rawson et al. (2002) is that they did not include a baseline reading condition in which readers only read the text. As a consequence, we do not know if comprehension was affected negatively or positively by concurrent metacomprehension judgments.

Factors contributing to metacomprehension accuracy

Another issue concerning JODs and POPs that we pursue in this study regards factors affecting metacomprehension accuracy. On the one hand, reading ability might be expected to strongly correlate with metacomprehension accuracy for both JOD and POP judgments. Such a result would align with results reported by Griffin et al. (2008), who found a positive relation between reading ability and relative metacomprehension accuracy. More skilled readers are generally more adept at manipulating and building mental representations during comprehension than less skilled readers (Snow 2002), and such representations likely utilize information, or cues, that align well with test performance. Hence, one hypothesis is that both JOD and POP accuracy should be positively correlated with reading ability.

Alternatively, one could hypothesize that JOD and POP accuracy will correlate with different factors. On the one hand, we specifically predict that POP accuracy will be significantly correlated with reading ability. This is because POPs are likely made by relying on a variety of information sources that are related to reading ability such as self-perception of their own ability level (e.g., Moore et al. 2005; Zhao and Linderholm 2008), domain familiarity (Glenberg and Epstein 1987; Glenberg et al. 1987; Maki and Serra 1992a), expectations about the nature of the questions (Maki et al. 1990), and heuristics acquired through experience (Linderholm et al. 2008). Reading ability measures capture some variance in these abilities because they are all related to the ability to understand text materials well.

On the other hand, we hypothesize that JOD metacomprehension accuracy should be more closely related to the extent to which participants use processing ease as a cue during the judgment task. Our reasoning is based on these two assumptions: 1) many JOD participants use processing ease as a cue; and 2) the experience of processing ease is diagnostic of the level of comprehension. In sum, whereas relative POP accuracy should be related to reading ability as measured by a standardized reading ability test, relative JOD accuracy should be more strongly correlated with a more specific measure representing participants' use of their experience of processing difficulty (or ease) as a cue during the metacomprehension judgments. We describe a specific measure to represent readers' use of processing difficulty as a cue in the results section when we report the analysis.

Domain knowledge is also recognized as an important factor in successful reading comprehension (Kintsch 1998). Although there is not much research on the specific relation between relative metacomprehension accuracy and domain knowledge, at least two accounts have been proposed previously that lead to an expectation for a generally positive relation between domain knowledge and metacomprehension accuracy. First, Wiley et al. (2005) proposed that metacomprehension would benefit from higher levels of domain knowledge because domain experts have a better sense of what it means to comprehend in that domain. Second, Griffin et al. (2008) claimed that higher-level domain knowledge frees up attentional resources for metacomprehension, leading to higher metacomprehension accuracy. Despite these theoretical expectations, recent work has suggested that domain knowledge may not contribute to a measure of relative metacomprehension accuracy, gamma, even though it contributes to absolute metacomprehension accuracy (Griffin et al. 2009). Thus, it is not clear whether there is a systematic relation between domain knowledge and relative accuracy.

Overview of the Study

We are interested in how different metacomprehension judgment tasks (JOD and POP) affect: 1) the information (cues) used to make metacomprehension judgments; 2) relative metacomprehension accuracy gamma; and 3) factors contributing to metacomprehension accuracy. In order to examine these issues, we presented an expository text on brain structure sentence-by-sentence, and randomly assigned participants to make either JOD or POP judgments. We simultaneously monitored participants' reading time for each sentence to measure moment-by-moment processing ease (or difficulty) that participants experienced. After the judgment task, participants answered multiple-choice comprehension questions about the text. Each comprehension question could be traced back to information contained in a sentence from the text. This allowed us to assess relative metacomprehension accuracy

(gamma). We also assessed general reading ability and domain knowledge to explore the relation among these individual difference factors with relative metacomprehension accuracy (gamma).

As a final remark in this section, we discuss the issue of ecological validity of asking participants to make sentence-by-sentence metacomprehension judgments in this study. One might expect this type of manipulation to alter reading processes. We expected that while the manipulation may affect reading processes, the extent of the influence would be relatively small, and the information gained by this method would outweigh this limitation. Our expectations were guided by the following findings. First, although it is not natural for readers to assign numerical ratings of their understanding after reading each sentence, it is known that individual sentences are an important meaningful unit that readers use to construct larger, more global understandings of the text. For example, the Construction-Integration model of reading comprehension emphasizes the importance of the proposition, which is similar to a clause or a simplest form of sentence, in the process of constructing a meaningful representation of text materials (Kintsch 1998). In addition, results from eye-tracking studies indicate sentence wrap-up effects, which suggest that a sentence is a meaningful unit of reading (e.g., Just and Carpenter 1980). Finally, readers often stop and regress to the beginning of a sentence when a sentence is difficult (Hyona et al. 2002) indicating that readers seem to monitor comprehension sentence-by-sentence thus, although readers do not naturally make explicit comprehension judgments after reading single sentence, it is natural for readers to implicitly evaluate their comprehension in such a manner. As such, making this process explicit should have only a small effect on how comprehension proceeds. We tested this assumption by comparing the judgment conditions to a read-only condition with respect to comprehension of the experimental passage.

Regarding the benefits of using such a methodology, one critical contribution of this study is to gain insight into some specific cues used in the course of moment-by-moment metacomprehension. In particular, one objective of this study was to examine the possibility that the experience of processing difficulty, as measured by reading time, would be one of the primary cues in JODs. As such, it is critically important to measure reading time and its contribution to metacomprehension judgments in the smallest grain size possible. As reported earlier, Rawson and Dunlosky (2002) failed to observe a significant correlation between reading time and judgment ratings when measuring reading time using a larger grain size – aggregated across subjects. As such, we followed the assumption that a sentence-by-sentence measurement of reading time and metacomprehension judgments would provide a better measure of the fine-grained moment-by-moment fluctuations of comprehension and metacognitive confidence.

Method

Participants

The participants in this study were 72 introductory psychology students who volunteered for course credit. Their mean age was 19.8 (SD=3.0). The participants were randomly assigned to the Read Only, Judgments of Sentence Difficulty (JOD), and Predictions of Performance (POP) conditions. Data from one participant in the POP condition and one participant in the JOD condition were discarded without analysis due to equipment failure.

Design, materials and measures

The independent variable in this study was judgment condition (JOD, POP, Read Only), which was manipulated between participants.

Text The text was an expository passage entitled "The Lower Level Brain Structures" obtained from a chapter on the brain in the psychology textbook "Psychology, Myers in Modules, sixth edition" (Myers 2001). The text was modified to increase variability in the difficulty of each sentence by eliminating and re-ordering sentences, replacing words, or modifying sentence structures. The final text was 1188 words long with 62 sentences of varying lengths ranging from 4 to 49 words per sentence.

Flesch reading ease score The Flesch reading ease score is a proxy measure for the normative difficulty level of the individual sentences used in this study. The score is computed based on the number of words and word length (the number of syllables) of each sentence. A sentence with a larger number of longer words will have a lower reading ease score, which means a more difficult sentence. This measure is assumed to be useful for representing the difficulty level of linguistic materials because word frequency and syntactic complexity are somewhat related to word length (i.e., longer words tend to have lower frequency) and sentence length (i.e., longer sentences tend to have a more complex syntactic structure) respectively. Although Flesch reading ease scores are typically used to represent the overall difficulty of a larger text than individual sentences, in the absence of a more "perfect" measure, we assumed that Flesch reading ease would provide a reasonable proxy measure of how difficult a sentence is relative to other sentences in an objective sense. The Flesch reading ease score for each sentence was calculated using Coh-Metrix (Graesser et al. 2004). The frequency distribution of reading ease levels is presented in Table 1.

Reading time Reading time of each individual sentence was measured to represent each individual reader's experience of processing difficulty of each sentence. Reading time refers to the time elapsed between the onset of the presentation of each sentence and when a reader presses the spacebar, which removes the sentence from the screen and advances to the next sentence. We divided reading time for each sentence by the number of words in the sentence to control for sentence length. This measure provides a sentence-by-sentence measure of a reader's subjective experience of processing difficulty.

Metacomprehension (JOD and POP) judgments For JODs, participants indicated their subjective difficulty estimate of the sentence they just read using a 4-point scale (1=very easy, 4=very difficult). For POPs, participants indicated their prediction of how well they thought they would answer a question based on the sentence they just read on a 4-point scale (1=likely to be wrong, 4=likely to be correct).

Table 1 Frequency distribution of sentences with different reading ease levels

Range of Flesch Reading Ease	Number of sentences
0 -19	6
20 -39	18
40 -59	23
60 -79	13
80 -100	2

Reading comprehension questions Fifty-nine 4-option multiple-choice comprehension questions were constructed based on the text to measure participants' comprehension of the sentences. Three types of questions were asked. The first type was 39 questions that probed information contained in the explicit content of individual sentences, called text-based questions. The second question type was comprised of 14 questions that required the reader to integrate information across multiple sentences, called bridging questions. The third type of question included 6 vocabulary questions that required participants to infer the meaning of difficult words from the text. The distinction between the three types of questions roughly follows the different levels of understanding postulated by the Construction Integration Model of comprehension (Kintsch 1988, 1998). Whereas text-based questions can be correctly answered based on text-based understanding (idea explicit) of individual sentences, correctly answering other two types of questions (i.e., bridging and vocabulary) requires a situation model level understanding which is a deeper understanding of the text by integrating across multiple sentences and/or background knowledge. When constructing the four answer options for each question, we followed a set of guideline to control the quality of distracters developed by Ozuru et al. (2007). Prior to data collection, for each question, a source sentence that contained the information needed to answer the question was identified for each of the 59 questions in order to subsequently compute gamma correlations. Identification of a source sentence is difficult and somewhat problematic for bridging questions because answering bridging questions requires integrating information distributed across two or more sentences. Hence, we first identified several (usually two) sentences that contained the idea relevant to a question, and then identified the one sentence that contained the most crucial idea to answer the question. For vocabulary questions, the sentence that contained the information relating to the word being queried was identified as the source sentence. These classifications were made separately by the first and second authors; disagreements were resolved by discussion. The questions in the comprehension test were ordered by their source sentences.

Confidence rating for reading comprehension questions Immediately after answering each comprehension question on the computer, participants gave a confidence rating for each response on a 4-point scale from 1 (not at all confident) to 4 (very confident).

General reading ability test Reading ability was measured by the Gates-MacGinitie Reading Test (GMRT; 3rd ed.), form S, grade level 10–12 (MacGinitie and MacGinitie 1989). The GMRT provides multiple short passages that vary in genre and difficulty. Test takers read the passage and answer multiple-choice questions about the text content by referring back to the source passage. The GMRT is understood to assess general reading ability in a broad stroke without providing a specific diagnosis on the strength and weakness of each individual.

Prior domain knowledge questions The prior domain knowledge questions consisted of 15 open-ended questions and 15 multiple-choice questions. Prior domain knowledge questions differed from comprehension questions. Whereas comprehension questions could be answered from information in the text and/or the reader's prior knowledge, prior domain knowledge questions could be answered solely using prior knowledge. That is, although prior knowledge questions queried information about brain or neuropsychology, prior knowledge questions asked information that were different from, and not directly related to, the information contained the text. More specifically prior knowledge questions were about MRI, cerebral cortex, action potential, neurotransmitter, etc. whereas the text and comprehension questions were about lower brain structures (e.g., medulla, thalamus, brain stem) and their functions.

Relative metacomprehension accuracy (gamma) Metacomprehension accuracy was represented by the gamma correlation between judgment ratings (JOD or POP rating) of the sentences and performance on individual comprehension questions (correct or incorrect) for each participant. The gamma correlation measure varies from −1 to +1, and represents the extent to which a reader's metacomprehension judgment covaries with actual comprehension performance. For example, if an individual estimates comprehension of sentence 1 to be higher than sentence 2, and the actual comprehension performance (e.g., as measured by performance on comprehension questions) is higher for sentence 1 than sentence 2, then the individual's relatively accuracy is high. The overall level of performance does not affect measures of relative metacognitive accuracy whereas absolute accuracy, such as overestimation, underestimation, and hit rate (i.e., whether the level of future performance score such as percentage correct can be accurately predicted), is affected by level of performance (Nelson 1984). In addition, relative accuracy is likely to be more closely relevant to the ability to appropriately distribute effort and time when studying a large amount of material because the judgment of how well one understands a specific topic (or section) relative to other topics (sections) would facilitate decision on how to distribute study time and effort.

Procedure

The experiment proceeded in four phases. In the first phase, participants were administered the GMRT with a 20-minute time limit. In phase two (i.e., the reading phase), participants read the brain structure text one sentence at a time, which was presented by E-prime (2000) on notebook computers. Reading was self-paced and each sentence was presented left justified on the computer screen. Participants pressed the spacebar to proceed through the text. For the Read Only group, when the spacebar was pressed, the sentence was removed and a new sentence appeared. For the prediction of performance (POP) group, after the spacebar was pressed the sentence was removed and a screen appeared that asked participants to predict how well they thought they would answer a question based on the sentence they just read on a 4-point scale (1=likely to be wrong, 4=likely to be correct). For the judgment of sentence difficulty group (JOD), after the spacebar was pressed the sentence was removed and a screen appeared that asked participants to indicate their subjective difficulty estimate of the sentence they just read using a 4-point scale (1=very easy, 4=very difficult). Both judgment groups made their rating by pressing the appropriate number on the keypad. After the rating was made, the next sentence appeared on the screen and the process continued until the text was finished. In all the three conditions, participants were informed, prior to reading, that they would answer comprehension questions based on the text once they were finished reading it in its entirety.

In the third phase, immediately after reading the text, participants answered the multiple-choice questions in the absence of the text. The questions were presented one at a time on a notebook computer using E-Prime (2000). They indicated their answer by pressing an appropriate key corresponding to one of the answer options. The Q, W, E, and R keys on the keyboard were relabeled the A, B, C and D keys for answer selection. Question answering was self-paced. After answering each question, participants gave a confidence rating from 1 (not at all confident) to 4 (very confident).

In the fourth phase, after the participants finished answering the comprehension questions, they were provided a booklet containing prior knowledge questions. The booklet was divided into three sections with a paper marking the division of the sections separating adjacent sections. First, participants answered 15 open-ended prior knowledge questions, and then answered 15 multiple-choice prior knowledge questions. Participants were

informed that once they moved to each new section (e.g., multiple-choice prior knowledge questions), they could not return to the previous section. Answering of the prior knowledge questions was self-paced. The prior knowledge measure was administered after the main task (judgment and comprehension question answering) to remove the possibility that exposure to the prior knowledge questions might influence reading and judgment behaviors (McNamara and Kintsch 1996). That is, given that the most important data in this study was participants' moment-by-moment reading behavior, we wanted to minimize participants' exposure to information relevant to the target text prior to the main task (i.e., JOD and POP). We were concerned that exposure to prior knowledge questions may make readers read the sentence faster due to priming and/or help readers think about the possible comprehension question, thereby influencing the metacomprehension judgment and its accuracy in an unintended way. In order to ensure that the reading comprehension task did not influence performance on the prior knowledge questions, we carefully verified that the prior knowledge questions could not be answered based on the text content. This procedure is in line with other published text comprehension studies (e.g., Ozuru et al. 2007; O'Reilly and McNamara 2007). In addition, even if reading text and/or answering comprehension question influenced (i.e. improved) performance on prior knowledge questions, that is better than prior knowledge questions influencing reading time, metacomprehension, and comprehension questions because they are the central measures of this study.

Results

There are three parts to the results section: 1) the cues affecting JOD and POP ratings, 2) metacomprehension (JOD and POP) accuracy (gamma), and 3) factors associated with JOD and POP accuracy. As a preliminary analysis, we computed the proportion correct on the comprehension questions and examined whether performance on this test differed across the judgment conditions to assess the possibility that administering the sentence-by-sentence metacomprehension task affected comprehension performance. A one-way ANOVA indicated no effect of condition on comprehension performance, $F(2, 67)=2.3$, $MSE=.018$, $p>.1$ ($M_{Read\text{-}Only}=0.57$, $SD=0.12$; $M_{POP}=0.49$, $SD=0.17$; $M_{JOD}=0.50$, $SD=0.12$). Thus, at least according to this analysis on comprehension performance, there is no evidence that asking participants to perform sentence-by-sentence JOD or POP judgments significantly affected their comprehension.

The cues affecting JOD and POP ratings

We examined for each individual participant whether JOD or POP ratings for each sentence correlated differently with two types of metacognitive cues. These two cues were each participant's sentence-by-sentence reading time per word and the Flesch Reading Ease score (Flesch 1948). We have two specific predictions: 1) Reading time will be more strongly correlated with JOD ratings than POPs because we hypothesize JODs, but not POPs, are primarily made based on the subjective experience of processing difficulty; and 2) the overall influence of these two cues (i.e., reading time and Flesch Reading Ease score) to judgment rating will be larger for JODs compared to POPs because POPs are assumed to be mostly based on factors other than processing difficulty. We suspect that POPs are much more complex in nature than JODs, involving processes such as activating and integrating one's own prior knowledge in the domain, analytic assessment of the text content, experience with similar questions, and so on.

In order to examine these predictions, it was necessary to assess the unique contributions of reading time and Flesch reading ease score to judgment ratings for each individual participant. Thus, we calculated, for each participant across items, two sets of partial correlations: 1) the partial correlation between judgment ratings and reading time per word, controlling for the effect of reading ease scores; and 2) the partial correlation between judgment ratings and Flesch reading ease scores, controlling for the effect of reading time.

In order to facilitate comparison between conditions, the signs of the correlations were appropriately adjusted (e.g., a positive correlation for the JOD condition and the POP condition both meant that slower reading time reflected less confidence in processing). Table 2 presents the partial correlations averaged across subjects separately for the two conditions. Zero-order correlations are also included in the Table 2 for reference. A 2 (type of correlation) X 2 (judgment condition) ANOVA yielded a main effect of condition, $F(1, 43)=28.9$, $MSE=.018$, $p<.001$, partial $\eta^2=.40$, and an interaction between judgment condition and type of correlation, $F(1, 43)=4.46$, $MSE=.023$, $p<.05$, partial $\eta^2=.09$. The main effect of judgment condition indicates that overall, correlations were higher in the JOD ($M=.18$, $SD=.10$) than in the POP ($M=.03$, $SD=.09$) condition. The interaction indicates that whereas JOD ratings correlate most strongly with reading time, POP ratings correlate most strongly with reading ease scores. However, the relations between reading ease scores and POP ratings are notably weak. The same analysis using the zero-order correlations yields a main effect of condition, $F(1, 43)=15.61$, $MSE=.029$, $p<.01$. The interaction between the condition and the type of correlation (cue) was not significant, $F(1, 43)=2.78$, $p=.1$. This latter result makes sense because zero-order correlations do not effectively isolate the unique contribution of two separate cues, hence, weakening the interaction.

As a follow up, we also performed sentence-based analyses on the metacomprehension (JOD and POP) ratings. In these analyses, we assessed whether reading time explained variance in judgment ratings over and above Flesch Reading Ease. First, separately for each condition, we computed sentence averages of judgment ratings across participants. Next, separately for each judgment condition, we performed a hierarchical linear regression predicting average metacomprehension ratings from the average reading time per word, and Flesch Reading Ease score, for each sentence. We entered reading ease scores in the first step and reading time per word in the second step, and assessed the change in R^2.

The results presented in Table 3 indicate that: 1) in the JOD condition, reading time explained unique variance in JOD ratings, suggesting that readers in part rely on their reading time as a cue when making JODs; and 2) almost no unique variance in POP ratings was explained by reading time.

Table 2 Average correlation across subjects between metacognitive ratings and reading time, and reading ease scores as a function of condition (standard deviations in parentheses)

	Average partial correlation between judgment and reading time (controlling for reading ease) Average Zero-order correlation between judg ment and reading time M (SD)	Average partial correlation between judgment and reading ease score (controlling for reading time) Average Zero-order correlation between judgment and reading ease score M (SD)
JOD	.22 (.15)	.15 (.12)
	.21 (.19)	.19 (.12)
POP	.00 (.15)	.07 (.17)
	.01 (.15)	.11 (.17)

Table 3 Regression analysis of the contribution of reading time and reading ease on metacognitive ratings

	R	R^2	ΔR^2	ΔF
JOD				
Reading Ease	.397	.158	.158	11.2**
Reading Time	.567	.321	.163	14.2**
POP				
Reading Ease	.250	.062	.062	4.0*
Reading Time	.263	.069	.007	0.4

Overall, all these analyses converge to support our two predictions: 1) the contribution of reading time is stronger for JODs than POPs; and 2) overall contribution of these two cues is smaller for POPs than for JODs. Finally, in addition to the above two findings, the results indicate a cross-over-interaction in which JODs are more strongly related to reading time whereas POPs are more strongly related to Flesch Reading Ease scores, suggesting that JODs and POPs rely on different cues.

Metacomprehension accuracy

We next assessed differences in metacomprehension accuracy between the two judgment conditions. We predicted POP accuracy to be higher than JOD accuracy because: 1) POPs are likely more transfer-appropriate to final testing than JODs; and 2) readers may take into consideration a retention estimate – how long one needs to remember information – for POPs but not for JODs (Maki and Serra 1992a; Rawson et al. 2002).

We computed gamma correlations between the JOD and POP ratings for each sentence and performance on the corresponding comprehension question (henceforth *reading phase gamma*). That is, we calculated a gamma correlation between the 39 text-based comprehension questions and the JODs or POPs for the 39 sentences that correspond with the question for each participant, and compared the gamma correlations based on these 39 items for each participant across conditions. We focused these analyses on gammas computed for the 39 text-based questions because the scoring of these items is most straightforward. Nonetheless, we also report gamma correlations for the 14 bridging inference and the 6 vocabulary questions to examine whether metacomprehension accuracy varies by question type.

Table 4 presents the gamma correlations separately for the three types of question as a function of judgment condition (reading phase gamma). We also included gamma correlations for the confidence ratings for the comprehension questions (henceforth *confidence gamma*). Confidence gammas were intended to provide a reference point of readers' ability to judge their comprehension because for the confidence ratings, participants could access their perceived difficulty in actually answering the question, rather than predicting it during the reading phase.

First, in order to examine the effect of condition on judgment accuracy, a one-way ANOVA on judgment accuracy (reading phase gamma) for text-based questions as a dependent variable and judgment condition as the independent variable was performed. The analysis indicated that the gamma correlations did not differ between the JOD (M=.09, SD=.39) and POP (M=.12, SD=.24) conditions, F<1.0. However, a one sample t-test comparing the mean gamma correlation of each group against zero indicated that the gamma was significantly different from zero only in the POP condition, $t(21)$=2.4, p<.05, Cohen's d=.50, hinting that judgment condition (i.e. JOD vs POP) may have affected relative accuracy, albeit to a moderate degree.

Table 4 Reading phase and confidence gamma correlations for three type of comprehension question as a function of judgment condition

	JOD		POP		Read only	
	Read Phase Gamma M (SD)	Confidence Gamma M (SD)	Read Phase Gamma M (SD)	Confidence Gamma M (SD)	Read Phase Gamma M (SD)	Confidence Gamma M (SD)
Textbased	.09 (.39)	.39 (.32)	.12 (.24)	.32 (.43)	NA	.48 (.29)
Bridging	-.04 (.64)	.39 (.47)	-.01 (.52)	.30 (.54)	NA	.41 (.52)
Vocab	-.25 (.78)	.17 (.90)	-.09 (.82)	.35 (.65)	NA	.22 (.80)

Turning to an additional question of how reading phase gammas vary as a function of question type, it is quite clear from Table 4 that the gammas for bridging and vocabulary inference questions were generally inaccurate across conditions. We conducted exploratory t-tests to assess differences between the gamma correlations for text-based questions and the other two types of questions. Overall, gammas for vocabulary questions (M=−.19, SD=.80) were significantly lower than gammas for text-based questions (M=.11, SD=.32), $t(38)=2.1$, $p<.05$, Cohen's d=.49. The difference between gammas for bridging questions (M=−.06, SD=.59) and the text-based questions was not significant, $t(43)=1.6$, $p<.1$, Cohen's d=.35. Note that all of the gammas for the vocabulary and bridging questions were near zero or negative. These values of negative gammas indicate that participants' performance on the comprehension questions tended to be worse when they provided higher metacomprehension ratings.

We also analyzed whether metacomprehension accuracy is better when measured in terms of participants' confidence ratings (i.e., confidence gamma) than in a predictive fashion as in the reading phase (i.e., reading phase gamma). Like the main analysis, we focused on metacomprehension accuracy for the text-based questions. A 2 (condition: JOD vs. POP) x 2 (phase: reading phase vs. confidence rating phase) ANOVA indicated a main effect of phase, $F(1, 43)=15.07$, MSE=.09, $p<.001$, *partial* η^2=.26, such that confidence gammas (M=.35, SD=.37) were significantly higher than reading phase gammas (M=.11 SD=.32). No other effects were significant. This result indicates that participants' ability to monitor whether they answered a question correctly is better than their ability to predict their future performance. This finding is generally in line with Maki and Serra (1992b), which reported an improvement in metacognitive accuracy when participants gained more information about their actual memory performance.

Overall, the analysis on the effect of task condition (JOD vs. POP) on metacomprehension accuracy (i.e., reading phase gamma) fails to provide clear evidence of the effect of judgment task (JOD vs POP) on relative metacomprehension accuracy even though POP accuracy was numerically higher. Additional analyses indicated that: 1) metacomprehension accuracy tended to be best for text-based questions; and 2) judgment accuracy was much better for confidence ratings than reading phase JOD or POP ratings.

Factors influencing metacomprehension accuracy

We assessed to what extent metacomprehension accuracy in the two conditions is correlated with various factors. Three specific factors were examined here: reading ability, prior knowledge, and readers' use of their experience of processing ease as a cue to make their

metacomprehension judgments. We predicted that: 1) POP accuracy would be most strongly correlated with reading ability, and 2) JOD accuracy would be more strongly correlated with readers' use of their experience of processing ease as a cue.

In order to examine our first prediction, we first correlated individual participants' metacomprehension accuracy (reading phase gammas for text-based questions) with reading ability and prior knowledge, separately for each condition. This analysis indicated that while POP accuracy correlated positively with reading ability (r=.533, p=.01), JOD accuracy was not significantly correlated with reading ability (r=.194, p>.40). Prior knowledge was not significantly correlated with metacomprehension accuracy across the JOD (r=−.165, p>.50) and POP (r=.220, p>.30) conditions. Hence, the analyses provide support for the first prediction, indicating that POP accuracy was significantly related to reading ability and JOD accuracy was not (even though there was a weak positive correlation between reading ability and JOD accuracy).

In order to examine the second prediction, we used the by-participant partial correlations between reading times per word and judgment ratings that were computed in the section Cues affecting JOD and POP ratings. We assume this measure reflects the extent to which readers effectively use their subjective experience of processing ease as a cue when making their metacomprehension judgments. Within each condition, we then computed correlations between this measure and metacomprehension accuracy. The analysis indicated that there was a positive correlation between readers' use of reading time (i.e., partial correlation between reading times per word and judgment rating) and JOD metacomprehension accuracy (r=.449, p=.03) but not POP accuracy (r=.048, p=.80).

As a follow up, we conducted a step-wise regression to further assess the relation between metacomprehension accuracy and the use of processing ease as a metacognitive cue. Separate regression models were conducted for the JOD and POP conditions. Each regression model used metacomprehension accuracy (i.e., reading phase gammas for text-based questions) as the dependent measure, and reading ability, prior knowledge scores, and the partial correlations representing readers' tendency to use processing difficulty from above as the predictors. For JOD accuracy, the stepwise regression model indicated that the best equation included only the effective use of processing difficulty as the predictor (b=.449, p<.05), F(1, 22)=5.30, MSE=.13, p<.05. The model explained 20 % of the variance in JOD accuracy. For the POP accuracy, the stepwise regression model indicated that the best equation included only the reading ability as the predictor (b=.533, p<.05), F(1, 21)=7.93, MSE=.04, p<.05. The model explained 28 % of the variance in POP accuracy. Overall, these analyses support our two predictions regarding factors affecting metacomprehension accuracy: 1) reading ability is a strong predictor for POP accuracy but not for JOD accuracy; and 2) effective use of processing ease (especially reading time) as a cue is a strong predictor for JOD accuracy but not POP accuracy.

Discussion

The goal of this study was to explore differences in how JODs and POPs are made in the context of science text comprehension when measured with a sentence-by-sentence reading task. We examined the effect of judgment condition on metacomprehension by examining: 1) the cues affecting JOD and POP judgments, 2) judgment accuracy gamma, and 3) factors related to JOD and POP accuracy. We evaluate each of these findings in turn.

Cues affecting JOD and POP ratings

Overall, the results of this study provide support for the notion that different types of information influence JODs and POPs. Our results suggest that JOD and POP judgments differ in the use of cues such that whereas JODs rely on the experience of processing ease or difficulty, POPs do not.

This finding contrasts with that of Rawson and Dunlosky (2002), who found that both ease of comprehension judgments (comparable to JODs in this study) and POPs were based on processing ease as defined by manipulation of text features. Our results differ from theirs in two ways. First, consistent with Maki et al. (1990), our results indicated that ease (difficulty) of comprehension judgments (JODs) were correlated with experience of processing ease (or difficulty) whereas POPs were not. Second, our proxy measure of processing difficulty, reading time, predicted judgment ratings in the JOD condition, whereas it did not in the study reported by Rawson and Dunlosky (2002). This difference may arise from various differences in methods. For example, Rawson and Dunlosky manipulated coherence relations of the sentences and letter deletion, whereas our manipulation was not limited to specific types of difficulty. Instead, we monitored natural gradients in sentence difficulty in a continuous manner using the Flesch Reading Ease measure.

In addition, the grain level of analyses is different between the two studies. Rawson and Dunlosky (2002) monitored reading time aggregated over multiple sentences. In contrast, our study monitored reading time on a sentence-by-sentence basis, which likely allowed for a more sensitive analysis of the relation between reading time and judgment ratings. Our findings are generally consistent with current theories of metacognition (e.g., Benjamin et al. 1998). That is, when metacomprehension judgments are based on the experience of processing difficulty, judgment ratings should be correlated with processing time (i.e., reading time), either because they are directly based on reading time or because they are based on factors affecting processing ease, which affects reading time. Our finding is also consistent with the text comprehension literature (Haberlandt and Graesser 1985), which indicates that increases in processing difficulty should result in increased reading time, which is what we observed.

Judgment accuracy

Turning to the effect of judgment condition on metacomprehension accuracy, two points are noteworthy. First, the gamma correlations (reading phase gammas for text-based questions) obtained in this study were notably lower than averages reported in previous work (Thiede et al. 2009). Second, we failed to observe a significant effect of judgment condition (JOD and POP) on the gamma correlations (Maki and Serra 1992a; Rawson et al. 2002). We believe these two issues are related. We discuss these two issues together here.

First, the gamma correlations in this study were lower than the general average of 0.27 as reported in previous reviews of the literature (Thiede et al. 2009). Given that relative accuracy of metacomprehension judgments is expected to increase when they are based on more specific units of text (e.g., sentence as opposed to paragraph; Dunlosky and Lipko 2007), this finding is rather surprising. Notably, however, we expected JOD accuracy to be relatively low because JODs are not likely transfer appropriate relative to the processes engaged when answering multiple-choice comprehension questions. Thus, the low JOD accuracies observed here are not surprising.

More surprising, however, is the low POP accuracy. Unfortunately, we do not have a definitive answer regarding why this occurred. Instead, we provide a speculation based on existing theories and our data. According to the cue-utilization framework,

metacomprehension accuracy should be affected by the extent to which the cues used to predict comprehension are diagnostic of actual comprehension performance (Thiede et al. 2009). This means, in order to selectively increase metacomprehension accuracy in one condition relative to others by manipulating judgment type, the manipulation needs to ensure that the diagnosticity of the cues used in one condition (POP) is significantly better than the other (JOD). In light of this theory, it is not entirely clear, in retrospect, whether the set up in this study facilitated the use of cues that are significantly more diagnostic of future performance on multiple-choice comprehension questions in the POP compared to the JOD condition. That is, although we successfully manipulated the use of cues across the JOD and POP conditions such that JOD but not POP participants used processing ease as a cue, it is not clear from our results what kinds of cues POP participants predominantly used when making their judgments.

Further, it is possible that the set up of the study did not render other diagnostic cues available or readily accessible to readers in the POP condition. For example, it may be difficult for participants to identify the key information in each individual sentence unless they have high levels of reading comprehension skill or prior knowledge. If so, the POP judgment task becomes challenging for many participants because their level of reading skill and/or prior knowledge may not be sufficient for deep analysis of individual sentence meaning. Indeed, POP accuracy was correlated with our measure of reading comprehension. This may signal a possibility that manipulating the metacomprehension task alone (see introduction section) may not be sufficient to substantially improve metacomprehension accuracy. Instead, the manipulation may need to involve changing the comprehension task (i.e., enriched encoding; Thomas and McDaniel 2007) that may lead participants to attend to the deeper meaning of the text which, in turn, would increase the saliency of more diagnostic cues in the process of metacomprehension (but see Maki and Serra 1992a; Rawson et al. 2002). In any event, these accounts are largely speculative; a combination of other factors such as the nature of the rating scale, text and measurement of comprehension may have also affected metacomprehension accuracy in complex ways. More research is needed to uncover the nature of how these attributes contribute to metacomprehension accuracy in greater detail.

In addition to the main findings discussed above, we also found that metacomprehension was inaccurate for predicting performance on so-called deeper questions, such as vocabulary and bridging questions. Regarding bridging questions, inaccurate metacomprehension is not surprising for several reasons. First, our instruction encouraged participants to judge their comprehension of individual sentences, whereas bridging questions require integration across sentences. Second, the low metacomprehension accuracy for bridging questions may be due, in part, to how we computed gamma for those items: we selected a single sentence that contained the most important information to that question, as opposed to a collection of the related sentences that contribute to the answer. Perhaps those gamma correlations would have been higher if we used multiple sources sentences instead. However, how to compute such a measure is not straightforward and would likely have its own unique limitations.

A more noteworthy finding was the low level of metacomprehension accuracy for vocabulary questions. This finding seems to suggest that: 1) most participants' comprehension of individual sentences remained relatively unfocused or undifferentiated, without including a representation of a very specific meaning of a key word that would afford discrimination of a target word from seductive distracters in the test phase as indicated by the poor (near chance level) answering performance on the vocabulary questions (M_{JOD}=.33, M_{POP}=.35,

M_{READ}=.37); and 2) most participants believed that this type of undifferentiated understanding was a sign of good comprehension as indicated by the negative metacomprehension accuracy gamma correlations. The finding that participants had a poor understanding and meta-understanding of words is concerning given that participants were explicitly directed to think about how much they understood each sentence on an individual basis. This suggests, indirectly, that having participants make sentence-by-sentence metacomprehension judgments does not improve comprehension.

Factors contributing to JOD and POP accuracy

The finding that reading ability was significantly correlated with POP accuracy is in line with the previous literature (e.g., Griffin et al. 2008). Given that reading ability was measured by a standardized reading ability test, which measures a variety of component skills involved in comprehension in a broad stroke, this finding is not surprising.

A new finding from this study was that JOD accuracy was more strongly correlated with readers' effective use of their experience of processing ease. The finding is important because the experience of processing ease may be diagnostic of future comprehension performance when making JODs during reading. However, this finding needs to be interpreted in the context of the low overall JOD accuracy observed in this study. That is, on the one hand, relying on processing ease when making JODs may not lead to accurate metacomprehension on average. Yet, readers who are particularly sensitive to processing ease while making JODs do tend to have more accurate metacomprehension than those making JODs who are not sensitive to processing ease.

It should be cautioned that even though we did not observe a significant correlation between the use of processing ease as a cue and POP accuracy, this does not necessarily indicate that people with higher POP accuracy are not sensitive to processing ease in general. As described earlier, we view the two different judgment tasks used in this experiment as a method to change how readers weight metacomprehension cues. The POP task is expected to induce participants to use other cues besides processing ease. By contrast, JOD judgments would be expected to enhance the saliency of processing ease cues for participants. An important direction for future research is to investigate other cues to which readers attend when making POP judgments.

Concluding Remarks

As a concluding remark, we mention contributions and limitations of the study. First, this study contributes to a better understanding of metacognition in general and metacomprehension in particular by providing a detailed moment-by-moment view of how two typical metacomprehension judgment tasks influence processes involved in metacomprehension. The results indicate that metacomprehension processes differ quite significantly between JOD and POP judgments. More specifically, depending on whether performing JODs or POPs, one may tend to use different cues, and this results in somewhat different metacomprehension accuracy. Finally, metacomprehension accuracy for JODs and POPs appear to be traceable to different factors such as reading ability or the way in which readers use reading time as a cue.

The study certainly has limitations regarding its generalizability to normal reading situations. By placing emphasis on metacomprehension judgments of individual sentences, the study may have altered participants' reading processes. To some extent we assessed the

impact of the judgment tasks by comparing comprehension accuracy to a condition in which the participants read the text without making judgments. Accordingly, the judgments had no effect on comprehension. This absence of an effect may be because readers engage in metacomprehension naturally. Nonetheless, engaging in explicit metacomprehension judgments is not naturalistic. Additionally, making metacomprehension judgments on a sentence-by-sentence basis may not align with typical metacomprehension monitoring during discourse comprehension, which may be better aligned with larger-than-sentence units. In fact, our observed low overall POP accuracy (see discussion on metacomprehension accuracy above) may, in part, be due to these atypical task demands. Thus, we acknowledge that some aspects of the results may be related to use of our specific research method, which forced readers to process the text in a very specific way.

That being said, however, the findings make important contributions to field of metacomprehension research by offering tractable avenues for future study. For example: What are the individual differences that underlie one's ability to assess processing difficulty? What cues do readers rely on when making POPs? What components of reading ability contribute to one's ability to accurately predict future performance? We hope that this study makes a significant contribution to the literature by beginning to delineate the types of processes engaged during metacomprehension and how different types of metacomprehension tasks relate to those processes.

References

Anderson, M. C. M., & Thiede, K. W. (2008). Why do delayed summaries improve metacomprehension accuracy? *Acta Psychologica, 128*, 110–118.

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.

Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: use of diagrams in computer-based training environments. *Instructional Science, 30*, 433–463.

Davey, B., & McBride, S. (1986). Effects of question-generation training on reading com-prehension. *Journal of Educational Psychology, 78*, 256–262.

Dunlosky, J., & Bjork, R. A. (Eds.). (2008). *A handbook of metamemoryand memory*. Hillsdale, NJ: Psychology Press.

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., & Hacker, D. (2002). Metacomprehension of science text: Investigating the levels of disruption hypothesis. In J. Otero, J. A. Leon, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 225–279). Mahwah, NJ: Lawrence Erlbaum Associates.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221–233.

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.

Graesser, A., McNamara, D. S., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers, 36*, 193–202.

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.

Griffin, T. D., Jee, B. D., & Wiley, J. (2009). The effect of domain knowledge on metacomprehneison accuracy. *Memory & Cognition, 37*, 1001–1013.

Haberlandt, K. F., & Graesser, A. C. (1985). Component processes in text comprehension and some of their interactions. *Journal of Experimental Psychology. General, 114*, 357–374.

Hyona, J., Lorch, R. F., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: evidence from eye fixation patterns. *Journal of Educational Psychology, 94*, 44–55.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review, 87*, 329–354.
Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review, 95*, 163–182.
Kintsch, W. (1998). *Comprehension: a paradigm for cognition*. New York: Lawrence Erlbaum Associates.
Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609–639.
Linderholm, T., Zao, Q., Therriault, D. J., & Cordell-McNulty, K. (2008). Metacomprehension effects situated within an anchoring and adjustment framework. *Metacognition and Learning, 3*, 175–188.
MacGinitie, W. H., & MacGinitie, R. K. (1989). *Gates-MacGinitie reading test* (3rd ed.). Itasca, IL: The Riverside Publishing Company.
Maki, R. H., & Serra, M. (1992a). Role of practice tests in the accuracy of test predictions on text material. *Journal of Educational Psychology, 84*, 200–210.
Maki, R. H., & Serra, M. (1992b). The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 116–126.
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, and Cognition, 16*, 609–616.
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.
Moore, D., Lin-Agler, L., & Zabrucky, K. M. (2005). A source of metacomprehension inaccuracy. *Reading Psychology, 26*, 251–265.
McNamara, D.S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes, 22*, 247–288.
Myers, D. G. (2001). *Psychology, sixth edition in modules*. New York: Worth Publishers.
Nelson, T. O. (1984). A comparison of current measures of accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109–133.
O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: good texts can be better for strategic, high-knowledge readers. *Discourse Processes, 43*, 121–152.
Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction, 25*, 399–438.
Rawson, K. A., & Dunlosky, J. (2002). Are predictions of performance for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 69–80.
Rawson, K., Dunlosky, J., & McDonald, S. (2002). Influences of metamemory on performance predictions for text. *Quarterly Journal of Experimental Psychology, 55A*, 505–524.
Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.
Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology, 28*, 129–160.
Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. S. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of Metacognition and Self-regulated learning*. Earlbaum: Mahwah, NJ.
Thomas, A. K., & McDaniel, M. A. (2007). Metacomprehension for educationally relevant materials: Dramatic effects of encoding-retrieval interactions. *Psychonomic Bulletin & Review, 14*, 212–218.
Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology, 132*, 408–428.
Zhao, Q., & Linderholm, T. (2008). Adult metacomprehension: Judgment processes and accuracy constraints. *Educational Psychology Review, 20*, 191–206.