

The rereading effect: Metacomprehension accuracy improves across reading trials

KATHERINE A. RAWSON and JOHN DUNLOSKY
University of North Carolina, Greensboro, North Carolina

and

KEITH W. THIEDE
University of Illinois, Chicago, Illinois

Guided by a hypothesis that integrates principles of monitoring from a cue-based framework of metacognitive judgments with assumptions about levels of text representation derived from theories of comprehension, we discovered that rereading improves metacomprehension accuracy. In Experiments 1 and 2, the participants read texts either once or twice, rated their comprehension for each text, and then were tested on the material. In both experiments, correlations between comprehension ratings and test scores were reliably greater for participants who reread texts than for participants who read texts only once. Furthermore, in contrast to the low levels of accuracy typically reported in the literature, rereading produced relatively high levels of accuracy, with the median gamma between ratings and test performance being $+.60$ across participants from both experiments. Our discussion focuses on two alternative hypotheses—that improved accuracy is an artifact of when judgments are collected or that it results from increased reliability of test performance—and on evidence that is inconsistent with these explanations for the rereading effect.

Accurately assessing one's text comprehension is instrumental in effectively learning new material. Accordingly, many researchers have investigated the theoretical bases of comprehension assessments, or *metacomprehension*, with one goal being to discover conditions that produce high levels of metacomprehension accuracy. In most metacomprehension research, participants read several short texts, rate their comprehension for each text, and then complete tests for each. With few exceptions, metacomprehension accuracy has been poor. For instance, Maki (1998b) reported that the mean correlation between ratings and test performance across 25 studies from her lab was only $.27$. On the basis of a review of the entire literature, Maki (1998b) concluded that "the low accuracy of text predictions may mean that students cannot predict performance well. Alternatively, low accuracy may indicate that our measurement of metacomprehension accuracy is too unreliable for us to detect changes" (p. 142).

Given such poor accuracy found in previous research, a question naturally arises: Can metacomprehension accuracy be improved? We provide an affirmative answer to this question by integrating current theory of metacognitive monitoring with theory of text comprehension.

Whereas previous metacomprehension research has often been grounded in theories of monitoring, it has typically not been informed by theory of comprehension (but see Otero & Kintsch, 1992; Weaver, Bryant, & Burns, 1995). On the basis of a theoretical link between monitoring and comprehension, we introduce a manipulation—rereading—that we predicted would improve metacomprehension accuracy. To this end, we first briefly describe a theory of monitoring that has guided some metacomprehension research. We then link this theory to processes underlying text comprehension to illustrate how rereading may improve metacomprehension accuracy.

According to current theories of metacognitive monitoring, a person's metacognitive judgments are inferential in nature. For instance, with respect to predicting subsequent memory for a recently studied word pair, an individual monitors various cues that are available during encoding of the item (Koriat, 1997), such as the fluency of encoding (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989). A judgment of learning for that item is then based on an inference about how those cues relate to performance (Schwartz, Benjamin, & Bjork, 1998). In the case of fluency, people judge that their memory for an item is better when it is fluently processed during study than when it is difficult to process (Begg et al., 1989). According to this framework, the accuracy of metacognitive judgments is a function of the degree to which the cues used in assessing comprehension empirically correlate with test performance (Koriat, 1993). More specifically, if the inferences are valid, accuracy will increase as the cues more highly correlate with subsequent test performance.

We thank Harry Bahrick, Charles Weaver, and Karen Zabrocky for helpful comments about this research. Experiment 2 was conducted with support by a grant from the McDonnell Foundation to K. T. Correspondence should be addressed to J. Dunlosky, P.O. Box 26164, Department of Psychology, University of North Carolina, Greensboro, NC, 27402-6164 (e-mail: dunlosky@uncg.edu).

Although these assumptions have been demonstrated with simple materials (e.g., general information questions and word pairs) and not all have been generalized to text, they are not controversial and arguably hold for metacomprehension judgments. For instance, it seems likely that metacomprehension judgments are inferential in nature (see, e.g., Maki, 1998b, for an inferential, cue-based approach to metacomprehension judgments). In the case of fluency as a cue for metacomprehension judgments, people presumably will judge that their understanding of a text is better when the text is fluently processed than when it is difficult to process. Moreover, metacomprehension accuracy will increase as the cues that judgments are based on correlate more highly with subsequent test performance.

One way to discover cues that will be highly predictive of performance on tests of comprehension is to understand the comprehension processes that provide many of those cues. The construction–integration model of comprehension (Kintsch, 1988) is amenable to this purpose. According to this model, various processes jointly construct a text representation that may include three levels—a lexical level, a textbase level, and a situation model (see Graesser, Millis, & Zwaan, 1997, for a discussion of other levels). At the lexical level, which contains the surface features of a text, words and syntactic information are encoded. At the textbase level, segments of the surface text are parsed into propositions (Just & Carpenter, 1992) and explicit links between text propositions are formed, which involves argument overlap and other factors (Britton & Gülgöz, 1991; Miller & Kintsch, 1980). The construction of the situation model also involves linking propositions; however, the linking of propositions here involves connecting text information with a reader's prior knowledge (McNamara, Kintsch, Songer, & Kintsch, 1996).

Construction of a text representation at all three levels proceeds in cycles in which the subprocesses operate on one text segment at a time. To the extent that all processes are appropriately executed, a complete and coherent text representation will be constructed. However, various factors can disrupt the success of these processes (Britton & Gülgöz, 1991; Just & Carpenter, 1992; Perfetti, 1993); such processing failures presumably provide important cues that inform comprehension assessments.

Many factors can produce processing failures. For example, readers lacking prior knowledge that is needed to successfully link text information will have difficulty processing text. Processing failures are also likely to occur when one is reading difficult expository texts, such as those typically used in metacomprehension studies. In many circumstances, the processes of text comprehension will not be successfully completed during one reading trial. Researchers have shown that comprehension improves across reading trials (Barnett & Seefeldt, 1989; Haenggi & Perfetti, 1992), which suggests that texts are not completely processed during the initial trial. Most

relevant here, Millis, Simon, and tenBroek (1998) investigated how text processing changes across reading trials. They examined the processing resources allocated to lexical access, textbase construction, and the construction of the situation model. During the first reading trial, processing was dominated by textbase construction. In contrast, when rereading occurred after a short delay, more resources were allocated to constructing a situation model (Millis et al., 1998).

When processing resources are dedicated to constructing a text base, such as during an initial reading, disruptions of processing at this level may serve as cues for individuals' assessing their comprehension. Specifically, comprehension assessments may be partly based on cues such as processing disruptions that occur while one is parsing propositions or establishing argument overlap between propositions. Basing one's assessments on disruptions that occur at this level of processing will constrain accuracy because they are not highly predictive of performance on tests that rely on the construction of a situation model (Kintsch, 1994). More important for the present research, however, are the cues that may be available when people reread texts. While one is rereading a text, many propositions that one has constructed during the first reading are available from the text representation (Jacoby, 1978). Thus, fewer processing disruptions will occur at the text base level, which in turn allows more resources for constructing a situation model (Millis et al., 1998). Such construction will be more or less successful across texts, depending on such factors as a reader's prior knowledge or the amount of global incoherence in the text (Kintsch, 1998). While individuals are rereading texts, varying levels of difficulty that they experience in constructing situation models serve as cues for comprehension ratings. Such cues will be more predictive of performance on comprehension tests because the situation model largely determines test performance (Kintsch, 1994; McNamara et al., 1996).

In summary, changes in comprehension processes may improve metacomprehension by providing cues that are predictive of comprehension. This account suggests that with rereading, comprehension processes will differentially inform comprehension ratings across trials, with the cues available during rereading being more predictive of test performance. Accordingly, a prediction is that accuracy will be higher after rereading texts than after a single reading. This prediction was empirically evaluated in the present research.

Just as important, the methods used across the present experiments provide evidence relevant to other hypotheses for why rereading may improve metacomprehension accuracy. For instance, a somewhat uninteresting hypothesis is that rereading serves to increase the reliability of test performance, which would in turn increase the likelihood that other factors could correlate with test performance. If rereading influences metacomprehension accuracy merely by its influence on the reliability of test

performance, we would also expect rereading to boost the accuracy of other kinds of metacognitive judgments, such as retrospective confidence judgments for test performance. To evaluate this possibility in Experiment 1, we collected these judgments to assess whether rereading improves the accuracy of comprehension ratings and retrospective confidence judgments. Other hypotheses are also evaluated, such as whether rereading improves metacomprehension by providing familiarity with all of the texts prior to one's making any ratings. Thus, we not only evaluate a critical prediction from the cue-based hypothesis, but evaluate other explanations for the possible influences of rereading on metacomprehension accuracy.

EXPERIMENT 1

Method

Participants, Design, and Materials. Eighty undergraduates from the University of North Carolina at Greensboro were randomly assigned to two groups (single reading or rereading). Seven texts (one practice text and six critical texts) were developed from expository passages appearing in a GRE preparation manual (Branson, Selub, & Solomon, 1987). Six five-alternative multiple-choice questions were developed for each text.¹ Each critical text covered one topic; the topic of each text is presented along with corresponding text characteristics in the Appendix. We used computers to control text presentation and collect data.

Procedure. All the participants received instructions and practiced the experimental tasks. The sample text was presented one sentence at a time (as in Maki, Jonas, & Kallod, 1994), beginning with the title. Each sentence remained on the screen until the participants pressed a button to advance to the next sentence. After the last sentence, a comprehension rating was prompted with the text title and the query, "How well do you think you will be able to answer a test question over this material in about 20 minutes? 0 (definitely won't be able), 20 (20% sure I will be able), 40 . . . , 60 . . . , 80 . . . , 100 (definitely will be able)." Next, the participants answered the sample questions. After answering each test question, they made a confidence judgment ("How confident are you that you answered correctly?") based on the 6-point predicted likelihood scale used for comprehension ratings.

For the critical trials, order for presentation of the texts was randomized, and the texts were presented as described above. The participants in the single reading group rated their comprehension for each text immediately after reading each one. After rating all texts, they completed the test questions and made a confidence judgment for each. The participants in the rereading group first read each text once. They then reread texts in the same order and immediately rated their comprehension after rereading each. After rating all texts, the participants completed the questions and made a confidence judgment for each.

Results and Discussion

All differences declared as reliable in both of the experiments have $p < .05$, and those declared as not reliable have $p > .10$. Although less relevant than predictive accuracy, we first present test performance and the magnitude of the ratings because restrictions in their range may constrain predictive accuracy. The proportion of correct test responses was marginally greater for the participants who reread texts ($M = .59$) than for those who read the texts once [$M = .52$; $t(78) = 1.89$, $SE_D = .038$, $p = .06$].

For the magnitude of the ratings, we computed the mean across individuals' mean ratings separately for both of the groups. The magnitudes of the ratings were not statistically different for those who reread text ($M = 49.9$) than for those who read text once [$M = 44.2$; $t(78) = 1.50$, $SE_D = 3.78$]. Thus, test performance and ratings were intermediate.

More important, however, is the influence of rereading on the accuracy of the participants' comprehension ratings. Metacomprehension accuracy was operationalized as the gamma correlation between an individual's comprehension ratings and test performance across texts (for a discussion of why gamma is the most appropriate measure of predictive accuracy for metacognitive judgments, see Gonzalez & Nelson, 1996; Nelson, 1984). The gamma correlation indicates the degree to which individuals accurately predict performance on questions for one text relative to the others. Note that these values are not single correlation coefficients computed across individual participants (such as those reported by Walczyk & Hall, 1989). Of the 80 participants, 5 were excluded from this analysis because of indeterminate correlations: 1 participant in the group reading texts once and 4 participants in the rereading group. As predicted, metacomprehension accuracy was reliably greater after rereading ($M \text{ gamma} = .57$) than after reading texts only once [$M = .24$; $t(73) = 2.58$, $SE_D = .13$], which we refer to as the *rereading effect*.

One hypothesis for the rereading effect is that rereading serves to increase the reliability of test performance, which in turn increases metacomprehension accuracy. In this case, the rereading effect would not be attributable to an actual influence of rereading on metacomprehension. If changes in reliability of test performance account for the rereading effect, we would also expect rereading to boost the accuracy of retrospective confidence judgments. To evaluate this hypothesis, we calculated the accuracy of confidence judgments for each participant by computing a gamma correlation between scores on the test questions and the corresponding confidence judgments. As with predictive accuracy described above, however, we first briefly present the magnitude of the confidence judgments for completeness. The mean across individuals' mean retrospective confidence judgments was 61.9 for those who reread text and was 53.1 for those who read text once [$t(78) = 2.45$, $SE_D = 4.44$], demonstrating intermediate levels of judgments for both groups.

Most important, the accuracy of retrospective confidence judgments was not reliably different for individuals who reread texts ($M = .47$) than for those who read texts once [$M = .50$; $t(78) = .58$, $SE_D = .06$]. Thus, the rereading effect does not extend to the accuracy of confidence judgments, which suggests that changes in reliability across groups cannot account for the rereading effect. Moreover, this dissociation implies that comprehension ratings and confidence judgments are based on different cues. Unlike comprehension ratings, which we propose are partly based on cues of disruptions from text process-

ing, Glenberg and Epstein (1987) suggest that confidence judgments are based on "specific experience gained from answering the inference" (p. 92), which often leads to relatively high levels of accuracy (e.g., Maki, 1998a; Maki et al., 1994).

EXPERIMENT 2

In metacomprehension research, any effects on accuracy can be as difficult to replicate as they are to produce. Therefore, Experiment 2 was a critical conceptual replication of Experiment 1, in which we evaluated whether the rereading effect would generalize to other texts and test materials. More important, one relatively uninteresting explanation for the rereading effect involves an individual's familiarity with all texts prior to making comprehension ratings. In particular, when the participants who read texts once made a comprehension rating, they were not familiar with texts that had not yet been rated, because the ratings were made immediately after each text had been read. By contrast, those who reread texts had a chance to familiarize themselves with each text before making ratings, which may in itself improve accuracy across trials. Recent evidence by Maki (1998a) is relevant to this hypothesis. Namely, she had some participants make ratings immediately after reading each text (as in the present Experiment 1) and had other participants make ratings only after reading all texts. Metacomprehension accuracy did not differ for these two groups, with the mean level of accuracy (as measured by gamma) for both groups being below .10. Even though this evidence indicates that familiarity with text alone does not produce high levels of accuracy, we evaluated this hypothesis for the rereading effect by having the participants in both groups read all the texts before making any ratings. If familiarity with the texts produces the higher level of accuracy demonstrated by the rereading group, metacomprehension accuracy will be equivalent for individuals who read texts once and for individuals who reread texts.

Method

Participants, Design, and Materials. Sixty undergraduates from the University of Illinois at Chicago were randomly assigned to two groups (single reading or rereading). The texts were nine expository passages used by Glenberg and Epstein (1985). The topic of each text is presented along with corresponding text characteristics in the Appendix. Six four-alternative multiple-choice questions were developed for each text. We used computers to control text presentation and collect data.

Procedure. For each participant, seven texts were chosen randomly from the pool of nine texts; six texts were randomly selected as the critical texts, and the remaining one was used for practice. All participants received instructions and practiced the experimental tasks. Texts were not presented sentence by sentence as in Experiment 1, but instead each text was individually presented on the computer screen. A practice text was presented for reading, and it remained on the screen until the participants pressed the return key. The participants then rated their comprehension of the text. The

comprehension rating was prompted with the title of the text at the top of the screen and the query, "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, the participants answered the questions written for the practice text.

For the critical trials, order of text presentation was randomized. The participants in the single reading group read the six texts first. After having read all six texts, they rated their comprehension for all texts. Finally, they answered test questions for each one. The participants in the rereading group read each of the six texts once and then reread the texts in the same order. After having reread all the texts, they rated their comprehension for all texts and subsequently answered test questions for each.

Results and Discussion

As in Experiment 1, the proportion of correct test responses was computed for each participant. As expected from previous research (e.g., Haenggi & Perfetti, 1992), mean performance across individuals was reliably greater for participants who reread texts ($M = .86$) than for those who read the texts once [$M = .70$; $t(58) = 5.25$, $SE_D = .03$]. We also computed the mean across individuals' mean ratings separately for each of the groups. The magnitude of the ratings were reliably greater for those who reread text ($M = 5.98$) than for those who read text once [$M = 5.12$; $t(58) = 3.53$, $SE_D = 0.25$]. Although test performance and ratings were relatively high, the measures were off the ceiling for both groups. Moreover, because these measures were highest for the rereading group, if such high performance constrains accuracy, it will tend to produce outcomes that are inconsistent with our prediction of a rereading effect.

Means across gamma correlations between each individual's ratings and performance were calculated as in Experiment 1. (Four participants in the rereading group were dropped from this analysis because of indeterminate correlations.) As predicted, the participants who reread texts were more accurate (M gamma = .55) than were those who read texts once [$M = .19$; $t(53) = 3.06$, $SE_D = .11$]. Thus, the rereading effect obtained in Experiment 1 was replicated with different texts, test materials, and procedures. Moreover, the results from Experiment 2 disconfirm the hypothesis that rereading improves metacomprehension accuracy by increasing familiarity with the texts. Whereas the participants in both groups were familiar with all six texts prior to making ratings of comprehension, relatively high levels of accuracy were evident only for the group in which the participants reread texts.

GENERAL DISCUSSION

The rereading effect was demonstrated across two experiments. Rereading produced a marked improvement in accuracy: Collapsed across experiments, 54 of 61 participants who reread texts had accuracy above chance, and median accuracy across all 61 participants was .60. Weaver and Bryant (1995) have published the only other research showing levels of accuracy above .50 (see Lin

& Zabrocky, 1998, for a comprehensive review of variables other than rereading that have been investigated in the metacomprehension literature).²

Why has metacomprehension accuracy often been so poor? As one explanation, Weaver and Bryant (1995, p. 18) offered that “the difficulty of texts . . . may have contributed substantially to the low levels of metamemory for text” reported by Glenberg and others. Although text difficulty may moderate metacomprehension accuracy, we used the difficult texts that have been frequently used in some previous studies of metacomprehension (*M* Flesch score = 45.5, in Experiment 2) as well as a set of texts that were presumably as difficult as these more standard texts (*M* Flesch score = 38.2, Experiment 1), suggesting that some factor(s) in addition to text difficulty contributed to the low accuracy previously reported. Also, in response to Maki’s (1998b) concern over methodological constraints as an explanation for low levels of metacomprehension accuracy (quoted above), the present results demonstrate that the methods typically used to evaluate metacomprehension accuracy can yield relatively high levels of accuracy under appropriate conditions.

We propose that these conditions can be identified if theory of monitoring is supplemented with theory of comprehension. Our present integration of these theories provides both an account of the rereading effect and testable predictions. According to this account, the rereading effect will not be obtained under several conditions. For instance, if readers can construct a situation model during an initial reading, cues from processing the situation model will be available both to individuals reading once and to those who reread. If the same cues are available during both readings, we would not expect differential levels of accuracy. This outcome may occur in various situations, such as when people are reading easy texts. In this case, however, accuracy will be low even after rereading, because disruptions in processing (which are likely to be minimal with easy texts) will not indicate differential comprehension across texts. Another prediction is that rereading will not improve accuracy when a comprehension rating is made for each section within a text about one topic. Construction of the situation model links information across sections, and hence cues from processing would be less specific to any one section. Thus, cues available during rereading may be more indicative of comprehension across the entire text rather than differential comprehension of individual sections.

Even though our specific theoretical account led to the discovery of the rereading effect, other factors presumably contribute to the observed levels of accuracy after rereading. Perhaps most important in establishing a newly discovered phenomenon, evidence from the present research disconfirms some hypotheses. First, an individual who rates comprehension during an initial reading typically will not have knowledge about the remaining texts. By contrast, individuals who reread can use their

knowledge of all texts when assessing comprehension. Thus, rereading may enhance accuracy by allowing a more informed comprehension rating. Although this explanation is intuitively plausible, the rereading effect occurred even when people rated comprehension after all texts had been read once (Experiment 2). Furthermore, as described in detail earlier, recent results from a different laboratory indicate that this factor will contribute little to the rereading effect (Maki, 1998a).

Second, rereading itself may influence the accuracy of comprehension ratings even though the ratings themselves are not influenced by rereading. For instance, rereading may increase the reliability of test performance, which in itself may enhance the correlation between ratings and performance. The lack of a rereading effect on the accuracy of retrospective confidence judgments is inconsistent with this hypothesis. Moreover, to further evaluate the influence of rereading on the correlation between comprehension ratings and test performance, we assessed accuracy in a follow-up group of 35 participants, who read all of the texts once, made comprehension ratings, read all texts again, and then received a test (materials and procedure were the same as those used in Experiment 1). If rereading enhances the correlation between ratings and comprehension without improving metacomprehension, accuracy for this group will be similar to that of the rereading groups. By contrast, the mean level of accuracy for this group was only .28, which is reliably different from the mean value of accuracy for the corresponding rereading group in Experiment 1 [$t(69) = 2.36, SE_D = .15$]. A key conclusion here is that the rereading effect is a function of assessing comprehension that occurs after rereading, and hence the effect is attributable to improving metacomprehension *per se*.

Although we believe that an integrative approach is well suited to discovery about metacomprehension, our specific theoretical instantiation likely underspecifies metacomprehension. Cues other than processing disruptions contribute to comprehension assessments (for a review, see Maki, 1998b) and may do so differentially across reading trials. During a single reading, disruptions that occur while one is constructing a textbase may even be dominated by other cues (Maki & Serra, 1992). Nonetheless, given the number and complexity of theories of comprehension and monitoring, a general approach of integrating such theories shows promise in providing foundations for developing instantiations other than the one described here, which can then be used to conduct progressive programs of research. Just as important, because the rereading effect is relatively large in magnitude, discovering conditions that (reliably) moderate the size of this effect seems viable, which in turn may be used to evaluate competing hypotheses.

In summary, we advocate an approach that integrates theory of metacognitive monitoring with theory of text comprehension. Explanations for the rereading effect other than the one described in the present article may

also be derived from this approach. These alternatives can be based on theories of comprehension and monitoring not considered here and will allow strong inferences to be made in developing metacomprehension theory. However, regardless of the specific interpretation of the present effects of rereading, it is currently one of the only manipulations that has been shown to consistently improve metacomprehension accuracy. Thus, besides having important implications for improving student scholarship, this new phenomenon provides a tool for guiding future research on metacomprehension.

REFERENCES

- BARNETT, J. E., & SEEFELDT, R. W. (1989). Repetitive reading and recall. *Journal of Reading Behavior*, **21**, 351-361.
- BEGG, I., DUFT, S., LALONDE, P., MELNICK, R., & SANVITO, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory & Language*, **28**, 610-632.
- BRANSON, M., SELUB, M., & SOLOMON, L. (1987). *How to prepare for the GRE*. San Diego: Harcourt Brace.
- BRITTON, B. K., & GÜLGÖZ, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology*, **83**, 329-345.
- GLENBERG, A. M., & EPSTEIN, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 702-718.
- GLENBERG, A. M., & EPSTEIN, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, **15**, 84-93.
- GONZALEZ, R., & NELSON, T. O. (1996). *Psychological Bulletin*, **119**, 159-165.
- GRAESSER, A. C., MILLIS, K. K., & ZWAAN, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, **48**, 163-189.
- HAENGGI, D., & PERFETTI, C. A. (1992). Individual differences in reprocessing of text. *Journal of Educational Psychology*, **84**, 182-192.
- JACOBY, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning & Verbal Behavior*, **17**, 649-667.
- JUST, M. A., & CARPENTER, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, **99**, 122-149.
- KINTSCH, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review*, **95**, 163-182.
- KINTSCH, W. (1994). Learning from text. *American Psychologist*, **49**, 294-303.
- KINTSCH, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- KORLAT, A. (1993). How do we know what we know? The accessibility model of the feeling of knowing. *Psychological Review*, **100**, 609-639.
- KORLAT, 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.
- LIN, L. M., & ZABRUCKY, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology*, **23**, 345-391.
- MAKI, R. H. (1998a). Predicting performance on text: Delayed versus immediate predictions and tests. *Memory & Cognition*, **26**, 959-964.
- MAKI, R. H. (1998b). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117-144). Hillsdale, NJ: Erlbaum.
- 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). The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 116-126.
- 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 & Instruction*, **14**, 1-43.
- MILLER, J. R., & KINTSCH, W. (1980). Readability and recall of short prose passages: A theoretical analysis. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 335-354.
- MILLIS, K. K., SIMON, S., & TENBROEK, N. S. (1998). Resource allocation during the rereading of scientific texts. *Memory & Cognition*, **26**, 232-246.
- NELSON, T. O. (1984). A comparison of current measures of feeling-of-knowing accuracy. *Psychological Bulletin*, **95**, 109-133.
- OTERO, J., & KINTSCH, W. (1992). Failures to detect contradictions in text: What readers believe vs. what they read. *Psychological Science*, **3**, 229-234.
- PERFETTI, C. A. (1993). Why inferences might be restricted. *Discourse Processes*, **16**, 181-192.
- SCHWARTZ, B. L., BENJAMIN, A. S., & BJORK, R. A. (1998). The inferential and experiential bases of metamemory. *Current Directions in Psychological Science*, **6**, 132-137.
- WALCZYK, J. J., & HALL, V. C. (1989). Effects of examples and embedded questions on the accuracy of comprehension self-assessments. *Journal of Educational Psychology*, **81**, 435-437.
- 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.
- WEAVER, C. A., III, BRYANT, D. S., & BURNS, K. D. (1995). Comprehension monitoring: Extensions of the Kintsch and van Dijk model. In C. A. Weaver III, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 177-193). Hillsdale, NJ: Erlbaum.

NOTES

1. The text and test questions were adapted from copyrighted material; for information on obtaining the materials, contact the corresponding author.
2. Walczyk and Hall (1989) have also been cited as demonstrating high levels of metacomprehension accuracy. However, their correlation of .58 was calculated between participants, as individuals only read and rated their comprehension for one text. Thus, although their study does provide evidence with respect to one kind of predictive accuracy, it did not assess the degree to which individuals can discriminate their level of understanding across texts and hence is less relevant to the current discussion.

APPENDIX A
Text Characteristics

Text	F reading ease	F-K level	No. Sentences	Words/Sentence
Experiment 1				
Obesity	44.3	12.6	23	26.1
Inventions	21.1	17.4	16	28.4
Intelligence	27.7	14.7	21	24.8
Guilt	46.4	11.2	17	21.1
Literature	47.6	11.4	25	22.4
Majority	41.9	12.6	16	23.7
<i>M (SD)</i>	38.2 (11.0)	13.3 (2.4)	19.7 (3.9)	24.4 (2.6)
Experiment 2				
Good hanging	44.0	11.6	12	18
Evolution	45.2	12.2	10	21.2
Blood sugar	45.0	11.3	6	23.2
Bose speakers	60.4	9.8	11	21.9
Genetics	32.6	13.2	9	20.2
Black holes	53.8	9.7	15	16
Borrowing	56.9	9.5	10	20.4
Sea levels	36.9	13.7	8	22.5
Viruses	34.7	15.2	6	27.3
<i>M (SD)</i>	45.5 (9.9)	11.8 (2.0)	9.7 (2.9)	21.2 (3.2)

Note—F, Flesch; F-K, Flesch-Kincaid; No. Sentences, number of sentences per text; Words/Sentence, number of words per sentence. *SD* = standard deviation for text characteristic within a given experiment.

(Manuscript received November 12, 1998;
revision accepted for publication September 23, 1999.)