Journal of Experimental Psychology:
Learning, Memory, and Cognition
2002, Vol. 28, No. 1, 69–80

0278-7393/02/$5.00 DOI: 10.1037//0278-7393.28.1.69

Are Performance Predictions for Text Based on Ease of Processing?

Katherine A. Rawson
University of Colorado at Boulder

John Dunlosky
University of North Carolina at Greensboro

In 4 experiments, the authors evaluated the hypothesis that performance predictions for text are based on ease of processing. In each experiment, participants read texts, predicted their performance for each one, and then were tested. Ease of processing was manipulated by having participants read texts that varied in coherence. Coherence was varied by manipulating causal relatedness across sentence pairs (Experiments 1 and 2) and by altering the structure of sentences within paragraphs (Experiment 3). In these experiments, prediction magnitudes increased as coherence increased, suggesting that predictions were based on processing ease. In Experiment 4, prediction magnitudes were greater for intact paragraphs than for paragraphs with letters deleted from some of the words. Discussion focuses on resolving apparent inconsistencies in the literature concerning whether processing ease influences performance predictions.

Metacomprehension, or the assessment of one's own understanding of text material, is an important aspect of learning from text. Metacomprehension arguably can have a central role in formal education and in many everyday contexts, such as learning from science text (Otero & Kintsch, 1992), narrative and expository texts (Weaver & Bryant, 1995), classroom lectures (Spires, 1993), and news stories (Schneider & Laurion, 1993), as well as reading from childhood (Ghatala, Levin, Foorman, & Pressley, 1989) to later adulthood (Lin, Zabrucky, & Moore, 1998). Accordingly, researchers have sought to understand how people assess their own comprehension (for recent reviews concerning different aspects of monitoring comprehension, see Hacker, 1998; Maki, 1998; and Otero, 1998).

A central question relevant to how people assess their comprehension is, What are the bases of metacomprehension judgments? One approach to understanding the bases of metacomprehension judgments derives from current metacognitive theory, such as the cue-utilization framework of metacognitive monitoring (Koriat, 1997). According to this framework, a person's metacognitive judgments are inferential and are in part based on beliefs about how various cues relate to subsequent test performance. The cues underlying metacognitive judgments may arise from aspects of the to-be-studied materials (Koriat, 1997; Weaver & Bryant, 1995), from the strategies used to encode items (Dunlosky & Nelson, 1994), or from retrieval attempts made at the time of judgment (Benjamin, Bjork, & Schwartz, 1998; Matvey, Dunlosky, & Guttentag, 2001; Morris, 1990).

Katherine A. Rawson, Department of Psychology, University of Colorado at Boulder; John Dunlosky, Department of Psychology, University of North Carolina at Greensboro.

Thanks go to Eileen Kintsch for helpful comments on an earlier version of this article; to D. Bryant, A. Glenberg, and R. Maki for insightful reviews; and to Erica Middleton for preparation and piloting of the materials for Experiment 4.

Correspondence concerning this article should be addressed to Katherine A. Rawson, Department of Psychology, University of Colorado, Campus Box 345, Boulder, Colorado 80309-0345. E-mail: krawson@psych.colorado.edu

Given the wide range of cues that may underlie various metacognitive judgments, an important challenge is to discover the cues that influence them for text. The present research was designed to examine one candidate cue, the ease of processing text material. Ease of processing has been shown to underlie metacognitive judgments for simpler materials such as individual words (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Matvey et al., 2001) and paired associates (Kidder, Dunlosky, & Hertzog, 1998). In Begg et al. (1989, Experiment 1), participants studied lists of words that varied in word frequency. One group of participants rated the memorability of each word, and another group rated the ease of studying each word. High-frequency words were rated as easier to study and were also judged to be more memorable than were low-frequency words. However, performance on the subsequent recognition test was less for high-frequency words than for low-frequency words. On the basis of the alignment of memorability ratings with ease-of-study ratings rather than with test performance, Begg et al. concluded that judgments of learning for words are based on ease of processing.

In extending the ease-of-processing hypothesis beyond simple verbal material to text, an important question arises: How might ease of processing be conceptualized with respect to reading text? Text comprehension involves many component processes, including encoding words and syntax, parsing this information into semantic propositions, establishing text-explicit links between propositions, making inferences, and integrating text information with relevant prior knowledge (Kintsch, 1998). Each of the component processes is susceptible to disruptions that may influence the ease of processing the text. For example, encoding words may be disrupted if the input is degraded, linking propositions may be disrupted if the propositions are distant from one another in the text, and inferring a connection between propositions may be disrupted if relevant prior knowledge is lacking. Of course, this summary of component processes and possible disruptions is not exhaustive. Rather, it is intended to illustrate a conceptualization of ease of processing as depending on disruptions of any process involved in reading texts.

Ease of processing has been implicated as a basis of metacognitive judgments for text. Rawson, Dunlosky, and Thiede (2000) proposed that metacomprehension judgments are sensitive to disruptions in the processing of the situation model. The situation model is a representation of the situation described by the text (Zwaan & Radvansky, 1998), involving integration of text information with prior knowledge (Kintsch, 1998). As performance on comprehension tests is largely dependent on the situation model (Kintsch, 1994), Rawson et al. (2000) proposed that judgments based on the ease of processing the situation model will typically be predictive of subsequent test performance. For instance, because rereading promotes processing at the level of the situation model (Millis, Simon, & tenBroek, 1998), a prediction is that metacomprehension judgments made after rereading will be sensitive to disruptions in situation-level processing and thus relatively accurate.

To test this prediction, Rawson et al. (2000) used a standard method from metacomprehension research. Participants read several texts, made a metacomprehension judgment for each text, and then completed comprehension tests for each. Half of the participants read each text twice prior to making judgments. As predicted, judgments made after rereading were more predictive of test performance than were judgments made after reading once. Although this effect of rereading is consistent with the hypothesis that judgments made after rereading are based on the ease of processing, Rawson et al. (2000) did not directly evaluate the hypothesized link between judgments and processing ease.

Evidence from other research suggests that processing ease may have a relatively minimal influence on performance predictions (Maki, Foley, Kajer, Thompson, & Willert, 1990; Maki & Serra, 1992). Maki and Serra (1992) had one group of participants predict future test performance after reading each text and had another group of participants rate the ease of comprehending each text. Maki and Serra hypothesized that if performance predictions were based on ease of comprehension, the two kinds of judgment would be equally predictive of performance. However, ease-of-comprehension ratings were significantly less predictive of test performance than performance predictions in two experiments, with a similar trend in a third experiment. These outcomes suggest that ease of processing plays a minimal role in predicting performance for text.

Both Maki and Serra (1992) and Rawson et al. (2000) relied on judgment accuracy to infer the relationship between processing ease and predictions, which has definite limitations. Predictive accuracy may be influenced by extraneous factors such as differential reliability of test performance across conditions (Schwartz & Metcalfe, 1994), reactive effects on performance from making judgments (cf. Spellman & Bjork, 1992), subtle changes in the query for the judgments (Kelemen, 2000; Rawson, Dunlosky, & McDonald, in press), and so on. Thus, differences in accuracy observed in previous research may implicate factors other than processing ease.

More direct evidence can be obtained by examining how manipulations of ease of processing affect judgment magnitudes (cf. Begg et al., 1987). Maki et al. (1990, Experiment 1) manipulated ease of processing by developing texts in which letters were deleted from some of the words in every other paragraph (McDaniel, 1984). One group of participants made a performance prediction after reading each paragraph of a text, whereas another group rated the ease of comprehending each paragraph. Ease-of-comprehension ratings were predicted to be greater for intact paragraphs than for paragraphs with deleted letters. On the basis of the hypothesis that performance predictions are based on ease of comprehension, predictions were also expected to be greater for intact paragraphs than for deleted-letter paragraphs. Ease-of-comprehension judgments were greater for intact paragraphs than for deleted-letter paragraphs, whereas the magnitudes of performance predictions for intact and deleted-letter paragraphs were not significantly different, suggesting that comprehension ease had a minimal influence on performance predictions.

Thus, although Rawson et al. (2000) argued that ease of processing is an important cue underlying performance predictions for text, existing research has failed to yield positive empirical support for the ease-of-processing hypothesis. Accordingly, we designed the present research to explore further the influence of ease of processing on performance predictions for text. To do so, we needed to manipulate the ease of processing text material. On the basis of the conceptualization of ease of processing, the critical manipulation in each experiment involved a theoretically motivated alteration of a text characteristic to induce disruptions in at least one of the component processes involved in reading text. In three of the current experiments, ease of processing was manipulated by varying coherence across text materials. Research on text comprehension has shown that texts are more difficult to process as the coherence of the material decreases (e.g., Britton & Gülgöz, 1991; Kintsch, 1998; Lorch & O'Brien, 1995; McNamara, Kintsch, Songer, & Kintsch, 1996; Myers, Shinjo, & Duffy, 1987). If performance predictions are based on ease of processing, then prediction magnitudes will increase as text coherence increases. In a fourth experiment, we manipulated processing ease by deleting letters from text, as in Maki et al. (1990), to permit reconciliation of present findings with those from previous research.

Experiment 1

One text characteristic that has been shown to influence the ease of processing is causal coherence, or the extent to which statements within a text are causally related. To the extent that causal links between ideas within a text are not explicitly indicated, readers engage in effortful, inferential processing to establish causal coherence (Fletcher, Hummel, & Marsolek, 1990; Myers et al., 1987).

Myers et al. (1987) developed materials in which pairs of sentences varied from low to high causal coherence, as indicated by normative ratings of causal relatedness. In Experiment 1, we used these materials to examine the relationship between ease of processing and magnitudes of performance predictions. As causal coherence decreases, ease of processing presumably decreases. On the basis of the hypothesis that ease of processing influences performance predictions, prediction magnitudes were predicted to decrease as causal coherence decreased.

Method

Participants and design. Eighty-one undergraduates enrolled at the University of North Carolina at Greensboro participated to partially satisfy a course requirement in introductory psychology. Level of causal coherence (high, moderate, or low) was a within-subject manipulation.

Materials. Materials included 24 sets of sentence pairs drawn from the materials developed by Myers et al. (1987). Each set included a second sentence, which described an outcome, and three possible first sentences, which had either high, moderate, or low causal relatedness to the second sentence. For example, the second sentence "He walked home soaking wet to change his clothes" would be preceded by either "Tony's friend suddenly pushed him into a pond" (high causal relatedness), "Tony had a fight with his friend near a pond" (moderate causal relatedness), or "Tony sat under a tree reading a good book" (low causal relatedness).

Procedure. Participants were instructed that they would read sentence pairs one sentence at a time, and that they would then make a performance prediction for each pair. Participants were informed that they would later be tested on their memory for each pair. Participants practiced each of the experimental tasks as described below for four sample pairs.

For the critical trials, each participant received eight high coherence pairs, eight moderate coherence pairs, and eight low coherence pairs. Across participants, all first sentences in each set were paired with the corresponding second sentence. Order of presentation of pairs was randomized anew for each participant. For each pair, the first sentence was presented for self-paced study. When the participant terminated study of the first sentence with a key press, the sentence was replaced with the second sentence for self-paced study. When the participant terminated study of the second sentence, it was replaced with the following prediction prompt: "You will be shown the first sentence from the pair you just studied in about 15 minutes. How confident are you that you will be able to recall the second sentence of the pair?" Participants made their predictions using a continuous scale from 0 (*definitely won't be able to recall*) to 100 (*definitely will be able to recall).* After making a prediction, the next pair was presented, and so on until all pairs had been studied and judged. Participants were then tested. For each pair, the first sentence appeared individually, and the participant was prompted to type in recall of the second sentence.

Results and Discussion

Judgment magnitudes. Given our theoretical aims, the magnitudes of the judgments were the primary measure of interest.[1] For each participant, the median prediction magnitude for low coherence pairs, moderate coherence pairs, and high coherence pairs was calculated. The mean across individual medians for each level of coherence is presented in Table 1. A repeated measures analysis of variance (ANOVA) revealed a significant effect of coherence, $F(2, 160) = 33.28$, $MSE = 4{,}249.43$.[2] Predictions were significantly less for low coherence pairs than for moderate coherence pairs, $t(80) = 5.46$, and for high coherence pairs, $t(80) = 6.87$. Predictions were also significantly less for moderate coherence pairs than for high coherence pairs, $t(80) = 3.10$. These results confirm the ease-of-processing hypothesis.

Reading times and test performance. Research on text comprehension has suggested that difficulties encoding text can increase reading time, reduce test performance, or both (Britton & Gülgöz, 1991; Kintsch & van Dijk, 1978). Accordingly, we examined each of these measures. For each participant, we computed the median reading time for the second sentence in low coherence pairs, moderate coherence pairs, and high coherence pairs (as in Myers et al., 1987). The mean reading time per sentence across individual medians for each coherence level is presented in Table 1. An ANOVA revealed a significant effect of coherence on reading time, $F(2, 160) = 6.52$, $MSE = 34.36$. Reading time was significantly greater for low coherence pairs than for moderate coherence pairs, $t(80) = 2.13$, and for high coherence pairs, $t(80) = 3.51$. Reading time for moderate and high coherence pairs did not differ significantly, $t(80) = 1.33$, but the overall trend is consistent with Myers et al.'s (1987) finding of decreasing reading time with increasing coherence. Although reading time is an indirect indicator of processing ease, these results support the assumption that ease of processing increased with causal coherence.

Table 1
Experiments 1 and 2: Judgment Magnitudes, Test Performance, and Reading Times as a Function of Level of Text Coherence and Kind of Judgment

	Coherence level		
	Low	Moderate	High
Experiment 1			
Performance predictions			
Judgment magnitude	40 (3)	49 (3)	54 (3)
Test performance	47 (3)	62 (3)	61 (3)
Reading time	7.82 (0.70)	6.95 (0.55)	6.55 (0.51)
Experiment 2			
Performance predictions			
Judgment magnitude	43 (5)	56 (5)	58 (5)
Test performance	31 (4)	46 (5)	50 (5)
Comprehension ratings			
Judgment magnitude	82 (4)	92 (3)	94 (2)
Test performance	30 (3)	42 (5)	44 (4)

Note. Coherence level refers to causal coherence between sentences in a pair. Values represent means across individuals. Performance is reported as percent correct. Reading time is presented as seconds of study for the second sentence in each pair. Standard errors of the mean are presented in parentheses.

To assess test performance, we propositionalized the second sentence of each set according to Bovair and Kieras (1985). The recall protocol of each participant was also propositionalized in the same manner. For each participant, the percentage of correctly recalled propositions for each sentence was calculated. A mean percentage correct was then calculated for low, moderate, and high coherence pairs. Mean percentage correct across individual means is presented in Table 1. An ANOVA yielded a significant effect of coherence on performance, $F(2, 158) = 30.65$, $MSE = 0.56$. (One participant was excluded from this analysis because of missing data.) Performance was significantly lower for low coherence pairs than for moderate coherence pairs, $t(79) = 6.19$, and high coherence pairs, $t(79) = 6.40$. Performance for moderate and high coherence pairs did not differ significantly, $t(80) = 0.20$.

Experiment 2

In Experiment 1, the magnitude of performance predictions increased monotonically with the causal coherence of sentence pairs. We interpreted this finding as indicating that processing ease was a basis for predicting performance, as text coherence presum-

[1] Although not a critical outcome for our present aims, we report judgment accuracy for this and subsequent experiments in Appendix A for interested readers.

[2] In all four experiments, all differences declared as significant had a value of $p < .05$.

ably influences the ease of processing text. Alternatively, people may have based their predictions on factors related to the output of rehearsal rather than on the ease of processing the texts.[3] Specifically, in anticipation of an upcoming recall test, participants may have used rehearsal strategies, such as by simply rehearsing the sentences or by attempting to form an image of the scenario described by each pair (cf. Dunlosky & Nelson, 1994). Their judgments may then have been based on factors relating to the output of the rehearsal strategies (e.g., the number of times the content was rehearsed, the quality of an image formed) rather than on the ease with which they read the text. Consistent with this hypothesis, reading times in our experiment were greater than those reported in Myers et al. (1987; reading times for second sentences reported by Myers et al. were 3.0–3.5 s, whereas they were 6.6–7.8 s in the present Experiment 1).

To rule out the possibility that magnitude differences were due to factors related to the output of strategic rehearsal, we conceptually replicated Experiment 1 using two groups in which participants were less likely to use a rehearsal strategy during study. To minimize the time participants would have available to use rehearsal strategies, the amount of time each sentence was presented was controlled. We used presentation rates that yielded study times in the same range as those reported in Myers et al. (1987) for both groups. In one group, participants made performance predictions. In the other group, participants studied sentences under incidental test instructions, and they rated their comprehension rather than predicting future test performance. As these participants did not anticipate a test, they would be even less likely to use rehearsal strategies. If factors relating to rehearsal were responsible for the pattern of judgment magnitudes in Experiment 1, then judgments in this experiment would not be expected to differ as a function of causal coherence for either group.

Method

Participants and design. Sixty undergraduates enrolled at the University of North Carolina at Greensboro participated to partially satisfy a course requirement in introductory psychology. Participants were randomly assigned to one of two judgment groups (performance predictions or comprehension ratings). Level of causal coherence was a within-subject manipulation.

Materials and procedure. Materials were the same as those used in Experiment 1. The procedure was the same as in Experiment 1, with the following exceptions: First, for all participants, each sentence was presented for 400 ms per word. This presentation rate resulted in a mean display time of 3.33 s for second sentences, which was similar to study times reported in Myers et al. (1987). Second, only those participants making performance predictions were informed that they would later be tested on their memory for each pair. Participants who made comprehension ratings were told that the purpose of the experiment was to explore how people assess their comprehension, and no mention of the subsequent test was made. After the offset of the second sentence for study, it was replaced with the following prompt for a comprehension rating: "How easy was it to understand the two sentences in the pair that you just read?" Participants made their comprehension ratings using the continuous 0–100 scale.

Results and Discussion

Judgment magnitudes. For each group, the mean across individual medians for each level of coherence is presented in Table 1. A 2 (kind of judgment) $\times$ 3 (level of coherence) mixed-factor ANOVA revealed significant effects of judgment group, $F(1, 56) = 53.00$, $MSE = 1{,}131.00$, and of coherence level, $F(2, 112) = 21.20$, $MSE = 142.50$. The interaction was not significant, $F < 1$. Judgment magnitudes were lower for performance predictions than for comprehension ratings, presumably because forgetting was factored into performance predictions in anticipation of the subsequent test (Rawson et al., in press). More important, the qualitative pattern of coherence effects was identical for comprehension ratings and performance predictions: Across groups, judgment magnitudes were significantly less for low coherence pairs than for moderate coherence pairs, $t(57) = 4.37$, and for high coherence pairs, $t(57) = 5.54$. Judgment magnitudes tended to be greater for high coherence pairs than for moderate coherence pairs, $t(57) = 3.10$, $p = .08$, although the trend did not reach significance. Overall, the pattern of magnitudes obtained under conditions in which individuals likely were not using rehearsal strategies during study is consistent with that found in Experiment 1—judgment magnitudes decreased monotonically as coherence decreased. Thus, even when the likelihood of rehearsal was minimized, the findings were consistent with the hypothesis that performance predictions are influenced by processing ease.

Test performance. Mean percentage correct across individual means for each coherence level is presented in Table 1. A 2 $\times$ 3 ANOVA was conducted, and the main effect of judgment group and the interaction were not significant, Fs < 1. Although one might have expected a main effect of judgment group favoring those who predicted performance (because they expected a test, whereas those making comprehension ratings did not), the lack of an effect is consistent with our assumption that neither group would explicitly rehearse the sentences given the relatively fast presentation rate. More important, a main effect for coherence level was obtained, $F(2, 106) = 22.10$, $MSE = 0.02$. Performance was significantly lower for low coherence pairs than for moderate coherence pairs, $t(55) = 4.37$, and for high coherence pairs, $t(55) = 5.54$. Performance for moderate coherence pairs and high coherence pairs did not differ significantly, $t(55) = 1.27$.

Experiment 3

In Experiments 1 and 2, we explored whether ease of processing is a basis for predicting performance using sentence pairs that varied in causal coherence. Simple text materials such as these are often used in comprehension research (Graesser, Millis, & Zwaan, 1997) to permit systematic control of text characteristics, and as such, can also be useful for metacomprehension research. Experiment 3 was designed to extend our findings to longer texts. In contrast to shorter texts, longer texts are more likely to differ from one another along multiple dimensions, such as text length or topic familiarity. These differences may provide other cues that dominate ease of processing when individuals predict performance. Consistent with this possibility, previous research suggesting that ease of processing has a minimal influence on performance predictions (Maki et al., 1990; Maki & Serra, 1992) used longer texts.

Our major goal of Experiment 3 was to evaluate the ease-of-processing hypothesis with respect to predicting performance for

[3] We thank an anonymous reviewer for suggesting this possibility.

longer text materials. In this experiment, we again used text coherence to manipulate ease of processing. Longer texts almost inevitably have some inherent incoherence, particularly the difficult, expository texts typically used in metacomprehension research. However, rather than selecting texts based on a subjective assessment of this inherent incoherence, we developed objective criteria based on the revision techniques of Britton and Gülgöz (1991) to systematically alter the coherence of our texts. Britton and Gülgöz developed theoretically motivated principles for improving text coherence based on the assumptions of the construction-integration model of text comprehension (Kintsch, 1988). Their revision techniques targeted characteristics of text that would make it difficult to construct a textbase level of representation (i.e., to link explicit text information across sentences). Of importance, they found that participants who read the revised version of a text were more efficient readers (i.e., recalled more propositions per minute of reading time) than participants who read the original text version, suggesting that the revised text was easier to process.

To manipulate ease of processing, we constructed texts that were either relatively coherent or relatively incoherent.[4] We first applied Britton and Gülgöz's (1991) techniques to repair incoherence inherent in the original versions of the texts, resulting in a coherent version of each text. For instance, we analyzed the order in which information was presented in every sentence of each text, examining whether the information was given (i.e., information that had been referred to in the previous sentence) or novel. We altered sentences as necessary to ensure that given information always preceded novel information within a sentence, to promote the ease of establishing connections between sentences. We then systematically altered each coherent version to construct an incoherent version for each text. For example, we rearranged each sentence so that new information preceded given information. These principled revisions decreased the ease of constructing a textbase representation.

Half of the texts read by each participant were coherent and half were incoherent (an example of each is presented in Appendix B). On the basis of the ease-of-processing hypothesis, prediction magnitudes were predicted to be higher for coherent texts than for incoherent texts.

Method

Participants and design. Twenty-five undergraduates from the University of North Carolina at Greensboro participated to partially satisfy a course requirement in introductory psychology. Text version (coherent or incoherent) was a within-subject manipulation.

Materials. Eight short expository texts were selected for adaptation from a Scholastic Assessment Test preparatory manual (The College Board, 1997). Across texts, the mean number of words per text was 370. Each text discussed a different topic, such as scientific discovery, Indian leadership traditions, and art forgery.

Two versions of each text were constructed—a coherent version and an incoherent version. Coherent versions were constructed on the basis of guidelines for principled revisions developed by Britton and Gülgöz (1991). We analyzed the coherence of each original text and repaired problems with referential coherence. First, for each key concept in a text, synonyms or ambiguous pronouns were replaced with the original term used to refer to the concept. Second, the presentation of information in each sentence was ordered such that given information (that which referred to or restated information contained in the previous sentence) preceded new information. These revisions were made to improve the ease with which explicit text information could be linked across sentences. For the resulting texts, Flesch readability scores ranged from 22.1–62.2 ($M = 44.3$), and Flesch–Kincaid grade-level scores ranged from 9.8–12.0 ($M = 11.6$).

The incoherent versions were constructed by making systematic modifications to the coherent version of each text. The first kind of modification was intended to increase the difficulty of establishing the connections between propositions explicit within each text. At every sentence boundary, argument overlap (i.e., the linking of two adjacent propositions by a common argument) was minimized by rearranging the clauses in each sentence, and where necessary, by rearranging the phrases within a clause. That is, for each sentence, the new information introduced by that sentence appeared at the beginning of the sentence, and any information referred to in the previous sentence appeared last. To increase the difficulty of parsing the explicit text information, syntactic modifications were made where possible without reversing the argument overlap modifications described above. Subject–verb orders were reversed, and prepositional phrases or object phrases were inserted between subjects and verbs. For the resulting texts, Flesch readability scores ranged from 24.0–61.4 ($M = 44.4$), and Flesch–Kincaid grade-level scores ranged from 10.0–12.0 ($M = 11.6$).

Eight 5-alternative multiple-choice questions were developed for each text. Four of the questions for each text tapped information that was explicitly stated in the text. The remaining four questions for each text tapped information that could be inferred from the text. Sample texts and test questions are included in Appendix B. Macintosh computers presented instructions, controlled text presentation and testing, and recorded all data.

Procedure. Participants were instructed that they would read texts one sentence at a time, and that they would then make a performance prediction for each one. Participants were informed that they would later be tested on their comprehension for each text. Participants practiced each of the experimental tasks as described below for a sample text and test questions.

For the eight critical texts, each participant was presented with four coherent versions and four incoherent versions. Across participants, each of the eight texts appeared in each version. The order of text presentation was randomized anew for each participant. Each text was presented one sentence at a time for self-paced study, beginning with the text title. Each sentence remained on the screen until the participant advanced to the next sentence with a key press (as in Maki, 1995; Maki, Jonas, & Kallod, 1994).

After reading the last sentence of a text, participants were prompted to make a performance prediction with the text title and the following query: "How likely is it that you will be able to correctly answer a test question about this text in about 20 minutes? 0 (*definitely won't be able*), 10 (*10% sure I will be able*), 20. . ., 30. . ., 40. . ., 100 (*definitely will be able*)."[5] The next text was then presented for study, and so on until all eight texts had been read and predictions had been made.

The test questions corresponding to the first text studied were then presented individually in random order. The five alternatives for each question appeared in random order. After answering each question, participants were prompted to make a postdiction with the text title and the following query: "How confident are you that you answered correctly? 0 (*don't know if I am correct*), 10 (*10% sure I am correct*), 20. . ., 30. . ., 40. . ., 100 (*definitely sure I am correct*)." The questions for the second text studied were then presented and rated in the same manner, and so on until the questions for each text had been completed.

---

[4] We greatly appreciate training provided by Dr. Bruce Britton in the use of these principled revision techniques.

[5] The scale used here is different than the continuous scale used in the first two experiments. This methodological difference between experiments is not critical because Lineweaver and Hertzog (1998) found that even when individuals make predictions on a continuous scale they tend to use the decade values (e.g., 0, 10, 20), and more important, because we are not making comparisons across experiments.

Results and Discussion

Judgment magnitudes. For each participant, the mean performance prediction for coherent texts and for incoherent texts was computed. The means across individual means for incoherent texts and for coherent texts are presented in Table 2. Predictions were significantly greater for coherent texts than for incoherent texts, $t(24) = 4.38$.

As predicted, performance predictions increased as text coherence increased, supporting the hypothesis that performance predictions for text material are based on processing ease. These findings conceptually replicate those from Experiments 1 and 2 and extend the conclusion that performance predictions are based on ease of processing for simple texts to longer text material.

Although the magnitudes of performance predictions were of primary interest, we also report judgment magnitudes for postdictions. As above, the mean postdiction magnitude for coherent and incoherent texts was computed for each participant. Across individual means, mean postdiction magnitude was 43.5 ($SEM = 4.6$) for incoherent texts and 48.1 ($SEM = 5.3$) for coherent texts, $t(24) = 2.14$, $p < .05$.

Reading times and test performance. For each participant, the median reading time per sentence for coherent texts and for incoherent texts was computed. Mean reading time per sentence across individual medians are presented in Table 2. Reading times were significantly faster for coherent texts than for incoherent texts, $t(24) = 4.09$.

For each participant, the mean percentage correct on questions for coherent texts and for incoherent texts was computed. Mean performance across participants is reported in Table 2. Performance scores did not significantly differ, $t(24) = 1.26$.

Experiment 4

The first three experiments each provided evidence consistent with the hypothesis that processing ease is a basis for performance predictions. Nonetheless, these findings contrast those reported by Maki et al. (1990). As described earlier, they found that ease-of-comprehension ratings were greater for intact text than for text with deleted letters, indicating that individuals were sensitive to differential ease of processing. However, performance predictions for intact texts and texts with deleted letters did not differ significantly, suggesting that processing ease had a minimal influence on performance predictions. What might explain the apparent inconsistency between this conclusion and the one derived from the findings reported in the current research?

One clue may be found in the patterns of test performance reported by Maki et al. (1990). Test scores were significantly greater for deleted-letter texts than for intact texts (.45 vs. .42, respectively), suggesting information from deleted-letter paragraphs was more readily retrieved than information from intact paragraphs. Text retrievability may influence performance predictions if individuals make a covert retrieval attempt at time of judgment and then base their predictions on factors related to the retrievability of the text. Consistent with this account, Morris (1990) reported that judgment magnitudes increased as the amount of text information retrieved in a speeded-recall trial increased. This retrievability hypothesis suggests that performance predictions for deleted-letter paragraphs would be higher than those for intact paragraphs, which Maki et al. (1990) disconfirmed. However, Maki et al. (1990) speculated that "subjects may have used some combination of ease of processing (which would produce higher ratings for intact text) and retrieval from the text representation (which would lower the ratings for intact text)" (pp. 614–615). Thus, the null effects reported by Maki et al. (1990) could have been due to offsetting influences of these two cues. As important, null effects would also be obtained if neither cue influenced performance predictions—it should be noted that processing ease plays no role in this case.

To reconcile our results with those reported by Maki et al. (1990), we conducted a replication and extension of their study. All participants read a text in which half of the paragraphs had letters deleted from some of the words and half of the paragraphs were left intact. Half of the participants rated the ease of comprehending each paragraph, whereas the other participants predicted performance for each paragraph. Half of the participants in each judgment group read texts in which the deleted-letter paragraphs were constructed as in Maki et al. (1990), which resulted in one to four letters being removed from deleted-letter words. The other participants in each judgment group read texts in which the deleted-letter paragraphs were constructed by removing all but one of the letters from deleted-letter words. We added this group to the design to reduce the performance advantage for deleted-letter texts, by decreasing the likelihood that deleted-letter words would be successfully completed. For ease of exposition, we will hereafter refer to these two groups as "easy completion" and "difficult completion," respectively. To foreshadow, we did obtain a group with equivalent performance for intact and deleted-letter paragraphs, which allowed a more definitive evaluation of the ease-of-processing hypothesis.

Table 2
Experiment 3: Judgment Magnitudes, Test Performance, and Reading Times as a Function of Kind of Text

	Kind of text	
	Incoherent	Coherent
Judgment magnitude	40 (3)	49 (3)
Test performance	41 (3)	45 (4)
Reading time	10.67 (0.51)	8.94 (0.65)

Note. Incoherent texts included revisions intended to increase the difficulty of constructing a textbase representation. Values represent means across individuals. Performance is reported as percent correct. Reading time is presented as seconds of study per sentence. Standard errors of the mean are presented in parentheses.

Method

Participants and design. One hundred twenty-eight undergraduates enrolled at the University of North Carolina at Greensboro participated to partially satisfy a course requirement in introductory psychology. Thirty-two participants were randomly assigned to each of four groups, defined by the factorial combination of two variables: kind of judgment (performance prediction or comprehension rating) and ease of completing deleted-letter words (easy or difficult completion). Kind of paragraph (deleted letters or intact) was a within-participant manipulation.

Materials. Base texts were the four *Science News* articles used by Maki et al. (1990). These expository articles discussed scientific issues

(e.g., health risks related to electrical fields) and were 21–27 paragraphs long. Flesch readability scores for the texts ranged from 22.8–42.0 (M = 32.2), and the Flesch–Kincaid grade-level score was 12.0 for each text. All texts were typed double-spaced. We replaced all deleted letters with an underlined space to indicate the number of letters missing from a word.

Experimental texts for the easy-completion group were constructed as described in Maki et al. (1990). Each article was divided into idea units, and letters were deleted from each word in every third idea unit for all paragraphs. First, all vowels were deleted, and then consonants were deleted as necessary to result in two deleted letters for words with four to five letters, three deleted letters for words with six to seven letters, and four deleted letters for words with eight or more letters. These texts were then piloted to ensure that words with deleted letters were still identifiable. Ten undergraduates wrote the missing letters into the words for each paragraph in each text. For words that were correctly completed by less than 40% of the pilot participants, two deleted letters were returned (or one if only two had originally been deleted). One deleted letter was returned to words that were correctly completed by 40%–70% of participants.

Experimental texts for the difficult-completion group were also divided into idea units, with letters deleted from the words in every third idea unit. In this group, all of the letters in the word were deleted except for one consonant.

In both the easy-completion and difficult-completion groups, two versions of each experimental text were developed. In each version, half of the paragraphs were intact (no letters deleted) and half of the paragraphs contained deleted-letter words. In one version, the even numbered paragraphs were intact, and in the other version the odd numbered paragraphs were intact. Each participant read one version of one text. For both groups, each version of each text was read an equal number of times across participants.

Test questions comparable to those used by Maki et al. (1990) were developed for each text. Two fill-in-the-blank questions were written for each paragraph, each requiring a one- to two-word answer.

Procedure. Participants were first given typewritten instructions. All participants were told that they would be reading a scientific text, and that some letters would be missing from some of the words. Participants were instructed to mentally complete these words as they read. All participants were also informed that after reading each paragraph in the text, they would be asked to make a judgment. In order to provide a replication of their original study, judgment prompts and the rating scale were identical to those used in Maki et al. (1990). Participants making performance predictions were told that for each paragraph they were to respond to the following prompt: "How certain are you that you will answer questions correctly over the above paragraph?" They were then shown the scale they were to use, rated from 1 (*not at all sure*) to 6 (*very sure*). Participants rating comprehension were told that for each paragraph they were to respond to the following prompt: "How easy was the above paragraph to understand?" They were shown the scale they were to use, rated from 1 (*not at all easy*) to 6 (*very easy*). All participants were told that they would be tested after they were finished reading the text, and the nature of the test was briefly described.

Participants were then given a text for self-paced study. The following judgment prompt appeared below each paragraph: "Please make your judgment for Paragraph #__," with the corresponding paragraph number filled in. A separate sheet displayed horizontal scales rated from 1 to 6 for each paragraph. Participants circled a value on the appropriate scale to indicate their judgment. After all paragraphs had been read and judgments had been made for each, the text and the rating sheet were removed and participants were given the test for self-paced completion.

Results and Discussion

Test performance. One of our goals in this experiment was to minimize performance differences between intact paragraphs and deleted-letter paragraphs, so as to reduce the potential influence of factors related to text retrievability on performance predictions. Accordingly, we first present test performance. As in Maki et al. (1990), each test answer was assigned a score from 0 to 2. Verbatim responses or close paraphrases were assigned 2 points. Partial answers or gist answers lacking sufficient detail received 1 point. Commissions or omissions received a 0. The point total for each participant was then converted to percent correct. For each group, the mean across individual percentages is reported in Table 3.

A 2 (kind of judgment) × 2 (ease of word completion) × 2(kind of paragraph) mixed-factor ANOVA revealed a significant main effect of ease of word completion, $F(1, 124) = 8.06$, $MSE = 0.03$. Overall, performance was greater for the easy-completion group than for the difficult-completion group. This main effect was qualified by a significant interaction of ease of word completion with kind of paragraph, $F(1, 124) = 5.51$, $MSE = 0.01$. Performance was greater for intact paragraphs than for deleted-letter paragraphs in the difficult-completion group, $t(63) = 2.28$. In contrast, performance for intact paragraphs and deleted-letter paragraphs did not differ significantly in the easy-completion group, $t(63) = 1.08$. No other main effects or interactions were significant, Fs < 1.

The statistically equivalent performance in the easy-completion group is apparently inconsistent with Maki et al.'s (1990) performance results. However, in their study, one group of participants wrote the missing letters into deleted-letter words whereas another group mentally filled them in (as in the present experiment). The performance advantage for deleted-letter texts was most pronounced when letters were written in and was minimal when letters were mentally filled in (as in the present experiment). Thus, the present findings are consistent with performance trends for the comparable group in Maki et al.'s (1990) study. Most important for present purposes, the statistically equivalent performance in this group permits evaluation of the ease-of-processing hypothesis.

Judgment magnitudes. For each participant, a median judgment for intact paragraphs and for deleted-letter paragraphs was calculated. For each group, the mean across individual medians is presented in Table 4.

A 2 × 2 × 2 mixed-factor ANOVA revealed a significant main effect for kind of judgment, $F(1, 124) = 6.50$, $MSE = 1.93$, with performance predictions being lower than comprehension ratings

Table 3
Experiment 4: Test Performance as a Function of Kind of Judgment, Ease of Completing Deleted-Letter Words, and Kind of Paragraph

	Comprehension ratings	Performance predictions
Easy completion		
Intact paragraph	25 (3)	23 (2)
Deleted-letter paragraph	27 (2)	25 (3)
Difficult completion		
Intact paragraph	20 (2)	22 (2)
Deleted-letter paragraph	17 (2)	17 (3)

Note. Performance is reported as mean percent correct across individuals. Standard errors of the mean are presented in parentheses.

Table 4
Experiment 4: Judgment Magnitudes as a Function of Kind of Judgment, Ease of Completing Deleted-Letter Words, and Kind of Paragraph

	Comprehension ratings	Performance predictions
Easy completion		
Intact paragraph	5.0 (0.2)	4.5 (0.2)
Deleted-letter paragraph	3.9 (0.2)	3.9 (0.3)
Difficult completion		
Intact paragraph	5.1 (0.2)	4.3 (0.2)
Deleted-letter paragraph	2.4 (0.2)	1.8 (0.2)

Note. Judgments were made on a 1–6 scale. Standard errors of the mean are presented in parentheses.

(cf. Experiment 2). As judgment group did not enter into any significant interactions, subsequent analyses were collapsed across kind of judgment. A significant main effect of ease of word completion was obtained, $F(1, 124) = 26.60$, $MSE = 1.93$. Judgments were lower for the difficult-completion group than for the easy-completion group. Most important, a significant main effect of kind of paragraph was also evident, $F(1, 124) = 153.60$, $MSE = 1.26$. Judgment magnitudes were lower for paragraphs with deleted letters than for intact paragraphs. This effect was significant in both completion-ease groups when examined separately: Participants made lower judgments for deleted-letter paragraphs than for intact paragraphs in the easy-completion group, $t(63) = 5.57$, and in the difficult-completion group, $t(63) = 11.20$. Finally, the interaction between ease of word completion and kind of paragraph was significant, $F(1, 124) = 38.60$, $MSE = 1.26$. Follow-up tests indicated that judgment magnitudes for intact paragraphs did not differ significantly for the easy-completion and difficult-completion groups, $t(126) = 0.11$. This outcome was expected, given that the intact paragraphs read by the two groups were identical. However, judgment magnitudes for deleted-letter paragraphs were lower for the difficult-completion group than for the easy-completion group, $t(126) = 7.71$. None of the other interactions were significant, Fs < 1.54.

As discussed above, the null effect of letter deletion on performance predictions reported by Maki et al. (1990) could have been due to offsetting influences of processing ease and text retrievability but could also indicate that neither cue influences predictions. The present findings disconfirm this latter interpretation. In particular, if neither cue influences performance predictions, predictions for intact and deleted-letter paragraphs would be expected to be equivalent for the difficult-completion group, despite reduced text retrievability (as indicated by test performance) and reduced processing ease (as indicated by lower comprehension ease ratings). However, judgments were lower for deleted-letter paragraphs than for intact paragraphs in this group.

Results from the difficult-completion group alone do not establish the role of processing ease in performance predictions, as lower predictions for deleted-letter paragraphs could have been due solely to differential retrievability. However, results from the easy-completion group demonstrate an influence of processing ease on performance predictions. For this group, test performance for intact and deleted-letter paragraphs was equivalent, essentially removing text retrievability as a basis for predictions. Nonetheless, predictions were still lower for deleted-letter paragraphs than for intact paragraphs—a difference that we attribute to differential ease of processing. This conclusion is further supported by the overall similarity between performance predictions and ease of comprehension ratings.

General Discussion

The present research was motivated by a question that is central to theories of metacognitive monitoring: What are the bases of metacomprehension judgments? We approached this issue from a cue-based account of metacognitive judgments, according to which metacomprehension judgments are based on inferences about how various cues relate to subsequent performance (Maki, 1998; Schwartz, Benjamin, & Bjork, 1997). The present research examined ease of processing as a cue for predicting performance, with four experiments yielding evidence confirming predictions from the ease-of-processing hypothesis: In all experiments, predictions increased as text coherence increased. This finding obtained with both short and long texts, and with manipulations along various dimensions of text coherence. Furthermore, patterns of performance predictions were highly similar to those for ease of comprehension ratings (Experiments 2 and 4), in which individuals are directly instructed to assess processing ease.

According to the conceptualization of processing ease forwarded in the introduction, ease of processing can result from any factor that disrupts a component process involved in reading texts. Consider the implications of the present findings within this conceptualization. In Experiments 1 and 2, our coherence manipulation targeted the ease with which inferences could be made. In Experiment 3, our coherence manipulation targeted the ease with which sentences could be parsed and links between propositions could be formed. In Experiment 4, the manipulation targeted the ease of lexical access. Although each of these manipulations primarily influenced processing at a different level of representation (the situation model, the textbase, and the surface level, respectively), performance predictions were influenced by each manipulation. Of course, people may not be sensitive to disruptions that occur in every kind of text processing, and hence performance predictions may not be influenced by all manipulations of processing ease. Accordingly, an important aim for future research will be to discover the degree to which disruptions from various sources differentially influence individual's performance predictions.

Although many results from the present research suggest that performance predictions for text are informed by ease of processing, other cues have also been implicated as bases for performance predictions for text (for a review, see Maki, 1998). Two cues in particular may have contributed to the present pattern of judgment magnitudes: text retrievability and reading time. We discuss each in turn.

As discussed in Experiment 4, factors related to text retrievability have been shown to influence performance predictions. Thus, another cue that may have produced the patterns of judgment magnitudes obtained in the present research is text retrievability, which was operationalized here and elsewhere (Maki et al., 1990) as test performance. Several findings are inconsistent with this interpretation of the present results. In Experiment 1, prediction

magnitudes increased with an increase from moderate to high causal coherence, whereas recall performance did not. A similar dissociation was obtained in Experiment 3: Performance predictions were significantly higher for coherent texts than for incoherent texts, whereas test performance for coherent and incoherent texts did not differ. Finally, performance predictions in Experiment 4 were lower for paragraphs with deleted-letter words than for intact paragraphs, even when performance was equivalent for the two kinds of paragraph. These outcomes are more consistent with the ease-of-processing hypothesis than with an account based on text retrievability.

Reading time may mediate the relation between ease of processing and performance predictions. Text coherence presumably affects ease of processing, which in turn can influence reading times (e.g., Myers et al., 1987). On one hand, performance predictions may be based on the reading times, as if an individual were explicitly considering how long it took to read a text when making a performance prediction for each text. This kind of mediator model has been recently supported in investigations of metacognitive judgments for relatively simple materials, such as single words and paired associates (e.g., Benjamin et al., 1998; Matvey et al., 2001). If reading time mediates the relation between predictions and differences in processing ease induced by text coherence, then a negative correlation will be observed between reading times and predictions. On the other hand, ease of processing may independently influence reading times and performance predictions, indicating that reading times do not mediate the relation. If so, we would not expect reading time to correlate with performance predictions.

To evaluate this possibility, we computed a gamma correlation for each participant between mean reading time (per sentence) for each text and the performance predictions for the two experiments from which reading time data were available (Experiments 1 and 3). The resulting mean across individuals' correlations was .05 in Experiment 1 and −.03 in Experiment 3. These results indicate that reading time had a minimal influence on performance predictions under two conditions, and hence rule out a model of metacomprehension judgments in which the relationship between processing ease and performance predictions is mediated solely by reading time.

Currently, the most parsimonious interpretation of the pattern of results across the present experiments is that performance predictions were informed by ease of processing. On the surface, this conclusion is apparently inconsistent with findings reported in other research (Maki et al., 1990; Maki & Serra, 1992). How can such inconsistencies be explained? On the basis of the finding that predictions more accurately predicted test performance than did ease of comprehension ratings, Maki and Serra (1992) concluded that performance predictions were not solely based on processing ease. Although we are not claiming that processing ease is the only cue that influences performance predictions, Maki and Serra's (1992) results are somewhat ambiguous with respect to the role of processing ease. In particular, predictive accuracy provides an indirect indication of the bases of metacomprehension judgments, as it is influenced by more than how people make judgments. For instance, the two groups in Maki and Serra's study made different pretest judgments (performance predictions or judgments of comprehension ease), and these judgments were differentially accurate. However, the two groups were also differentially accurate when prompted to make the same kind of posttest judgments (confidence in having answered test questions correctly). Postdictions from the performance prediction group were significantly more accurate than postdictions from the comprehension ease group. Because postdiction prompts were identical for both groups, differences in accuracy were not expected. Thus, the difference in postdictive accuracy may reflect some other (unspecified) difference between groups that may also have influenced differences in predictive accuracy.

Rather than judgment accuracy, judgment magnitudes were the primary measure of interest in the present research and in Maki et al. (1990, Experiment 1). In both studies, text characteristics were manipulated to induce differential ease of processing. In contrast to the present research, Maki et al. (1990) found that performance predictions did not vary with the differential ease of processing induced by the text manipulation (letter deletion), suggesting ease of processing may minimally influence predictions. However, Maki et al. (1990) suggested that the predictions may have been based on cues that exerted opposite influences on the magnitudes of predictions, resulting in a null effect of the deleted-letter manipulation on performance predictions. As discussed in detail above, our replication and extension of their study is consistent with this idea.

Jacoby and Brooks (1984) insightfully noted that

> when talking to students, one gains the impression that they use their lack of 'stumbling' as a criterion for judging their comprehension of material. If they can read through a chapter without being forced to backtrack, they are convinced that they understand the material and are then outraged when they do poorly on a classroom examination covering that material. (p. 39)

Thus, although the idea that processing ease can affect metacomprehension is not new, the present research provides some of the first empirical evidence that ease of processing influences people's predictions of performance for text.

References

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28,* 610–632.

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.

Bovair, S., & Kieras, D. E. (1985). A guide to propositional analysis for research on technical prose. In B. K. Britton & J. B. Black (Eds.), *Understanding expository text* (pp. 315–362). Hillsdale, NJ: Erlbaum.

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.

The College Board (1997). *10 real SATs*. New York: College Entrance Examination Board.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language, 33,* 545–565.

Fletcher, C. R., Hummel, J. E., & Marsolek, C. J. (1990). Causality and the allocation of attention during comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 233–240.

Ghatala, E. S., Levin, J. R., Foorman, B. R., & Pressley, M. (1989). Improving children's regulation of their reading PREP time. *Contemporary Educational Psychology, 1,* 49–66.

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

Hacker, D. J. (1998). Self-regulated comprehension during normal reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 165–191). Hillsdale, NJ: Erlbaum.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. *The Psychology of Learning and Motivation, 18,* 1–47.

Kelemen, W. L. (2000). Metamemory cues and monitoring accuracy: Judging what you know and what you will know. *Journal of Educational Psychology, 92,* 800–810.

Kidder, D. P., Dunlosky, J., & Hertzog, C. (1998, November). *Fluency effects on judgments of learning and quality of encoding ratings*. Paper presented at the 39th Annual Meeting of the Psychonomic Society, Dallas, TX.

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*. Cambridge, England: Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363–394.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126,* 349–370.

Lin, L. M., Zabrucky, K., & Moore, D. (1998, March). *Effects of text difficulty on calibration of comprehension*. Poster presented at the 44th Annual Meeting of the Southeastern Psychological Association, Mobile, AL.

Lineweaver, T. T. & Hertzog, C. (1998). Adults' efficacy and control beliefs regarding memory and aging: Separating general from personal beliefs. *Aging, Neuropsychology, and Cognition, 5,* 264–296.

Lorch, R. F., & O'Brien, E. J. (Eds.). (1995). *Sources of coherence in reading*. Hillsdale, NJ: Erlbaum.

Maki, R. H. (1995). Accuracy of metacomprehension judgments for questions of varying importance levels. *American Journal of Psychology, 108,* 327–344.

Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117–144). Hillsdale, NJ: Erlbaum.

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., 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, and Cognition, 18,* 116–126.

Matvey, G., Dunlosky, J., & Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (JOLs): An analytic or nonanalytic basis for JOLs? *Memory & Cognition, 29,* 222–233.

McDaniel, M. A. (1984). The role of elaborative and schema processes in story memory. *Memory & Cognition, 12,* 46–51.

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

Millis, K. K., Simon, S., & tenBroek, N. S. (1998). Resource allocation during the rereading of scientific texts. *Memory & Cognition, 26,* 232–246.

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

Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language, 26,* 453–465.

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

Otero, J. (1998). Influence of knowledge activation and context on comprehension monitoring of science text. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 145–164). Hillsdale, NJ: Erlbaum.

Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science, 3,* 229–235.

Rawson, K. A., Dunlosky, J., & McDonald, S. L. (in press). Influences of metamemory on performance predictions for text. *Quarterly Journal of Experimental Psychology*.

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

Schneider, S. L., & Laurion, S. K. (1993). Do we know what we've learned from listening to the news? *Memory & Cognition, 21,* 198–209.

Schwartz, B. L., Benjamin, A. S., & Bjork, R. A. (1997). The inferential and experiential bases of metamemory. *Current Directions in Psychological Science, 6,* 132–137.

Schwartz, B. L., & Metcalfe, J. (1994). Methodological problems and pitfalls in the study of human metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 93–113). Cambridge, MA: MIT Press.

Spellman, B. L., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science, 3,* 315–316.

Spires, H. A. (1993). Learning from a lecture: Effects of comprehension monitoring. *Reading Research and Instruction, 32,* 19–30.

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.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123,* 162–185.

Appendix A

Judgment Accuracy

For each participant, a gamma correlation between judgments and test performance across texts was computed (for rationale on using gamma as a measure of judgment accuracy, see Nelson, 1984). For each participant, a separate gamma correlation was also computed within each coherence level. Means across individual correlations for each experiment are presented in Table A1.

Table A1
Predictive Accuracy as a Function of Kind of Judgment, Experiment, and Level of Text Coherence

	Kind of judgment	
	Performance predictions	Comprehension ratings
Experiment 1		
Low coherence	.28 (.05)	—
Moderate coherence	.13 (.06)	—
High coherence	.10 (.06)	—
Across coherence levels	.24 (.03)	—
Experiment 2		
Low coherence	.43 (.09)	.18 (.09)
Moderate coherence	.17 (.08)	.45 (.13)
High coherence	.30 (.09)	.20 (.13)
Across coherence levels	.35 (.06)	.37 (.06)
Experiment 3		
Incoherent texts	.38 (.13)	—
Coherent texts	.13 (.17)	—
Across coherence levels	.31 (.08)	—
Experiment 4		
Easy-completion group		
Deleted-letter paragraphs	.08 (.09)	.24 (.09)
Intact paragraphs	.13 (.12)	.02 (.13)
Difficult-completion group		
Deleted-letter paragraphs	−.03 (.13)	.22 (.10)
Intact paragraphs	.07 (.13)	−.19 (.11)

Note. Values represent means across individuals. Standard errors of the mean are presented in parentheses. Dashes indicate that no data were obtained.

In Experiment 1, an analysis of variance (ANOVA) revealed a marginal main effect for coherence level, $F(2, 146) = 3.02$, $MSE = 0.72$, $p = .05$. Given the omnibus test approached significance, we report the follow-up tests for interested readers. Namely, performance predictions were significantly more accurate for low coherence texts than those for moderate coherence texts, $t(73) = 2.08$, and high coherence texts, $t(76) = 2.25$. Accuracy did not differ significantly for moderate coherence and high coherence texts, $t(74) = 0.14$.

In Experiment 2, a 2 (kind of judgment) × 3 (level of coherence) mixed-factor ANOVA yielded no significant main effects but a significant interaction, $F(2, 78) = 5.69$, $MSE = 0.15$. For performance predictions, accuracy was marginally greater for low coherence texts than for moderate coherence texts, $t(24) = 1.91$, $p < .07$. The accuracy of performance predictions did not differ significantly for low coherence versus high coherence texts, $t(23) = 0.65$, or for moderate coherence versus high coherence texts, $t(24) = 1.20$. For comprehension ratings, accuracy was significantly greater for moderate coherence texts than for high coherence texts, $t(16) = 3.23$. However, the accuracy of comprehension ratings did not differ significantly for low coherence versus moderate coherence texts, $t(17) = 0.90$, or for low coherence versus high coherence texts, $t(18) = 0.05$.

In Experiment 3, accuracy of performance predictions for incoherent and coherent texts did not differ significantly, $t(16) = 0.60$. For this experiment, gamma correlations between postdictions and performance were also calculated. Across individual correlations, the mean correlation was .31, $SEM = .05$.

In Experiment 4, a 2 (kind of judgment) × 2 (ease of word completion) × 2 (kind of paragraph) mixed-factor ANOVA did not reveal a main effect of ease of word completion, $F(1, 90) = 2.99$, $MSE = 0.31$, $p < .09$, but the interaction between kind of judgment and kind of paragraph was significant, $F(1, 90) = 5.67$, $MSE = 0.36$. Judgment accuracy for intact paragraphs and for deleted-letter paragraphs did not differ for participants who made performance predictions, $t(48) = 0.91$. In contrast, judgment accuracy was greater for deleted-letter paragraphs than for intact paragraphs for participants who rated comprehension, $t(44) = 2.46$. No other main effects or interactions approached significance, Fs < 1.22.

In summary, the modest levels of predictive accuracy obtained were similar to those reported in most previous research (Maki, 1998). Maki et al. (1990) reported that performance predictions were more accurate for deleted-letter paragraphs than for intact paragraphs. Consistent with this finding, we found greater accuracy for low coherence texts in Experiment 1 and a similar (but nonsignificant) trend in Experiment 3. However, results from Experiment 2 were inconclusive, and Experiment 4 failed to replicate their finding.

(*Appendixes continue*)

Appendix B

Sample Text Versions and Test Questions for Experiment 3

Coherent Text Version

Television newscasts have allowed viewers to form their own opinions about various political events and political leaders by relaying information and images instantly. In many instances, television newscasts have even fostered active dissent from established government policies. For example, in the 1960s, it is no coincidence that the civil rights movement took hold in the United States with the advent of television, as television newscasts were able to convey both factual information and such visceral elements as outrage and determination. Only when nightly television newscasts showed the civil dissent occurring in places like Selma and Montgomery to all of America's viewers did the issue of civil rights become a national concern rather than a series of isolated local events. Television newscasts relayed reports from cities involved to an entire nation of viewers, showing viewers the scope of the dissent and informing the dissenters that they were not alone. The ability of television newscasts to foster dissent has also been affected by increasingly widespread access to personal video cameras, as the news presented on television newscasts now comes from personal videos as well as from professional news agencies. Personal video cameras have been used by dissenters across the world to gather visual evidence of human rights abuses. Visual evidence gathered by personal video camera has then been transmitted across otherwise closed borders by television newscasts. The personal video camera is claimed by Jack Nachbar, a professor of popular culture, to be a "truth-telling device that can cut through lies." Nachbar's claim presumes, though, that the television viewer can believe what he or she sees. But the videotape that appears on television newscasts can, like still photography, be staged and even faked, so the motivation of the photographer must be taken into account. If photographers who are propagandists for some government utilize computer-generated effects, viewers will have more trouble believing what they see. However, even if seeing television newscasts is not automatically believing, at least seeing is seeing—and seeing is the fastest road to freedom in some repressive governments.

Incoherent Text Version

Viewers have been allowed to form their own opinions by television newscasts about various political events and political leaders that relay information and images instantly. Active dissent has even been fostered from established government policies by newscasts in many instances. In the 1960s, it is no coincidence for example that the civil rights movement took hold in the United States with the advent of television, as able to convey both factual information and such visceral elements as outrage and determination were newscasts. Rather than a series of isolated local events, the issue of civil rights became a national concern only when nightly television newscasts showed the civil dissent occurring in places like Selma and Montgomery to all of America's viewers. Informing the dissenters that they were not alone, television newscasts showing viewers the scope of the dissent relayed reports from cities involved to an entire nation of viewers. As the news presented on television newscasts now comes from personal videos as well as from professional news agencies, the ability of television newscasts to foster dissent has also been affected by increasingly widespread access to personal video cameras. Across the world, visual evidence of human rights abuses has been gathered by personal video cameras used by dissenters. Then transmitted across otherwise closed borders by television newscasts is the visual evidence gathered by personal video cameras. Jack Nachbar claims as a professor of popular culture that the personal video camera is a "truth-telling device that can cut through lies." That the television viewer can believe what he or she sees is presumed by Nachbar's claim, though. But into account must be taken the motivation of the photographer, as like still photography, videotape that appears on television newscasts can be staged and even faked. If propagandists for some government are photographers who utilize computer-generated effects, the more trouble believing what they see viewers will have. Even if seeing television newscasts is not automatically believing, at least seeing is seeing, however—and the fastest road to freedom is in some repressive governments seeing.

Inferential Test Question

The passage is primarily concerned with ways in which

A. television affects viewers by its presentation of news (correct)
B. television newscasts deliberately distort information
C. truth frustrates efforts by the media to constrain it
D. viewers of television newscasts cannot sort out fact from fiction
E. governments manage to control television newscasts

Explicit Test Question

The author explicitly states that the believability of television news may be compromised by

A. effects produced by computers (correct)
B. videos from personal cameras
C. photographers for professional news agencies
D. established government policies
E. reports that are transmitted across closed borders

Received August 29, 2000
Revision received April 18, 2001
Accepted May 6, 2001 ■