BEING REALLY, REALLY CERTAIN YOU KNOW THE MAIN IDEA DOESN'T MEAN YOU DO[1]

Michael Pressley
University of Maryland

Elizabeth Ghatala
University of Houston

Jennifer Pirie and Vera E. Woloshyn
University of Western Ontario

An important assumption in most models of skilled reading is that readers monitor their comprehension of text (e.g., Baker & Brown, 1984), that they are aware of when content is understood and when understanding is less than complete. Such awareness is presumed to play a critical role in regulating comprehension processes. Thus, if a reader believes prose is being encoded and interpreted as intended by an author, there is no reason to modify processing of text. If text is being read quickly, beginning at the first word of every paragraph and proceeding to the end of each paragraph, it is likely that rapid beginning-to-ending reading will continue. Alternatively, feelings of miscomprehension can direct rereading of material already covered or alter processing of subsequent content (e.g., cause the reader to read more slowly and carefully). In short, comprehension monitoring has been conceptualized as a critical executive process in skilled reading, regulating other processes that affect comprehension.

Much of the early work on comprehension monitoring was done within the error detection paradigm. Students read text containing inconsistencies or errors (e.g., two statements in a story about fish, one claiming they live where there is no light and the other that fish select their food by color). If readers noticed such problems, the argument was that they were monitoring their comprehension (e.g., Markman, 1977, 1979)—they were detecting that their understanding of one part of a text conflicted with what they understood another part of it to mean. For the example, if the subject coded that fish lived in a completely dark environment, the statement about selecting food on the basis of color should be surprising and result in a report of text inconsistency. Unfortunately, there were alternative interpretations of failures to report errors

in text, ones making obvious the inadequacy of the error detection approach as an index of comprehension monitoring. Three of the possibilities mentioned by Winograd and Johnston (1982) were that failures to report errors could be due to lack of prior knowledge, reflect readers' general belief that printed texts do not contain errors, or result from rationalizations made by readers to explain away inconsistencies.

The interpretive difficulties with error detection stimulated the development of alternative methods for measuring comprehension monitoring. One technique developed in our laboratories (Pressley, Ghatala, Woloshyn, & Pirie, in press) has been to ask people main idea questions about text they have just completed reading (e.g., What is the author's purpose in this passage? What would be a good title for the passage?) and to elicit ratings of confidence in their responses. If a reader is monitoring comprehension, then confidence should be high when responses to main idea questions are adequate and low when responses are inadequate. Our most important finding to date, however, is that adult readers usually are moderately confident about the correctness of their answers to main idea questions, regardless of the adequacy of their responses (Pressley et al., in press, Experiment 2). Most striking, they are overconfident about poor answers, causing them to bypass opportunities to restudy text (Pressley et al., in press, Experiment 1). In short, adults do not seem to monitor well their comprehension of main points in text, a type of monitoring failure that can undermine executive actions (e g , deciding to reread) that potentially could improve understanding of text.

In the study reported here, we reexamined comprehension monitoring using the main idea-question paradigm The particular problem studied here was whether comprehension monitoring (and thus, responses to main idea questions) might be more adequate if students were induced to use an exceptionally stringent criterion, one more exacting than their usual standard Thus, students in a *high-certainty* condition read stories accompanied by main idea questions. They were instructed to continue reading and processing the passage and its accompanying question until they could respond to the question with a high degree of confidence. In contrast, *one-reading* subjects were asked simply to read the passage one time, to provide an answer to the passage question, and to rate the certainty of their answer. Based on Pressley et al. (in press, Experiment 2), the expectation was that one-reading subjects would rate both their correct and incorrect answers approximately equally and about 5 on a 7-point scale (i e confident, although not extremely confident). That is, their confidence following one reading was expected to be well below the 7-point ceiling of the scale, and thus, there would be room for the high-certainty instruction to increase confidence.

One well established finding in the error detection paradigm is that instructions to shift criteria affect performance, in general, any information provided to subjects about what constitutes an error increases accuracy in reporting errors consistent with the criteria specified in the instructions (e.g Baker, 1985, Elliott-Faust & Pressley, 1986; Markman & Gorin, 1981). Thus, we surmised that asking subjects to use a different criterion than the one they normally would adopt (viz., one higher than their usual one) might increase critical evaluation of their first responses. If so, they might review text additionally to determine if their first attempt to summarize the text theme really produced an answer veridical with the meaning in the prose. If this manipulation

was successful in increasing the quality of answers provided to main idea questions, it would suggest a simple intervention for improving monitoring of main idea comprehension. Readers could be encouraged to adopt especially stringent criteria for deciding they have understood the most important idea in a passage.

METHOD

Subjects

Forty undergraduates (22 females, 18 males; mean age = 19.5 yrs; age range = 18 to 25 yrs) who were enrolled in a first-year university course served as subjects in the experiment. Subjects were randomly assigned either to the high-certainty condition or the one-reading condition.

Materials

Subjects read 10 passages (a different random order for each participant), each between 200 and 500 words in length. These were taken from SAT verbal subtests (e.g., College Entrance Examination Board, 1988) and covered literary, scientific, and social scientific topics. The following example is typical of the length and difficulty of these readings:

> As soon as cable service was restored after the earthquake, Baron Okura replied to architect Frank Lloyd Wright's inquiry with a message of congratulation:
> HOTEL STANDS UNDAMAGED AS MONUMENT OF YOUR GENIUS.
> HUNDREDS OF HOMELESS PROVIDED FOR BY PERFECTLY MAIN-
> TAINED SERVICE. CONGRATULATIONS, OKURA.
> Never one to display undue reticence in such matters, Wright speedily convened a press conference at which he said nothing to dissuade reporters from drawing the inference that the Imperial Hotel was the only building in Tokyo that had remained standing through the disaster. In fact, however, hundreds of other solid masonry buildings in both Tokyo and Yokohama also withstood the quake—most notably those of British architect Josiah Condor, whose numerous structures suffered considerably less damage than Wright's. Nonetheless, the Imperial Hotel's thoroughly undeserved fame as the only building that had stood up through the great Tokyo quake was to prove far more unshakable than the edifice itself; and Wright's renown as the man who had designed and built it flourished accordingly. While by no means wholly responsible for the architectural revolution that was to revitalize the world's cities during the next four decades, the worldwide repute of Wright's Imperial Hotel was to facilitate and hasten its progress. By the time this famous edifice was demolished in 1967, the great earthquake had been instrumental in altering not only the appearance of Tokyo but also that of many of the other great cities in the world (College Entrance Examination Board, 1984. p. 55).

Each passage was accompanied by a short-answer question tapping the overall theme of the passage. Subjects were asked either the main idea of the passage, its primary purpose, what the author principally wanted to discuss, or for a title summarizing the passage content. Thus, for the example passage subjects were asked to complete the sentence, "The primary purpose of the passage is to . . . ."

After completing the 10 passages and questions accompanying them, the subjects

took a complete 40-item SAT verbal section. This provided an estimate of individual differences in verbal competence.

Procedure

Before a passage was presented, the subject read the question accompanying it. High-certainty subjects then read the passage under an instruction to continue reading until they could provide an answer to the question with a very high degree of certainty (i.e., they could give an answer they were "*very, very* sure of." In contrast, one-reading subjects were presented the question before reading and were told to read it from beginning to end one time only and to generate an answer based on the single reading. After producing an answer, subjects in both conditions rated their confidence of correctness on a 1 (absolutely sure answer is incorrect) to 7 (absolutely certain answer is correct) scale, with the midpoint of 4 correponding to "50/50 chance the answer is correct" Following the rating, the subject proceeded to the next passage.

RESULTS AND DISCUSSION

Verbal Ability

High-certainty subjects averaged 24.10 items correct ($SD = 4.42$) out of 40 on the SAT verbal section, the corresponding figure for one-reading subjects was 23.05 ($SD = 4$ 47) These means did not differ significantly, $t(38) = 0.75$, $p > .50$, suggesting approximately equal verbal ability in the two conditions. This was as expected since there was random assignment of participants to the two conditions, that it was so, however, makes less likely that other significant differences between conditions that occurred could be explained away as artifacts of differences in ability between the two conditions.

Reading Time

High-certainty subjects spent more time reading the passages than one-reading participants; the total reading time was 27.25 mins. ($SD = 12.88$ mins.) in the high-certainty condition versus 18.83 mins. ($SD = 2.72$ mins.) in the one-reading condition, $t(38) = 3$ 03, $p < 01$ (Kirk, 1982, for this and all subsequent statistical references).

Performance on the Main Idea Questions

All answers that addressed the question and were consistent with the text were considered correct, with two raters achieving 95% agreement, disagreements were resolved by discussion For instance, for the question about the purpose of the example passage, any answer referring to how the earthquake was responsible for shifting world architecture in the ensuing years was accepted as correct.

The extra time in the high certainty condition did not translate into significantly better performance on the main idea questions. High-certainty subjects averaged 5.70 correct out of 10 compared to 5.00 in the one-reading condition, $t(38) = 1.02$, $p > 90$($MS_E = 4$ 689). That is, even the students instructed to use a high criterion

provided errant respoises more than 40% of the time. The proportions of subjects answering a question correctly did not differ significantly for 9 of the 10 passages, greatest $X^2(1) = 1.62$, $p > .05$ for these nine passage questions. For one passage, more high-certainty subjects (11 of 20) responded appropriately to the question than one-reading participants (2 of 20), $X^2(1) = 9.23$, $p < .01$.

Certainty Ratings

Despite objective performance far below ceiling, high-certainty subjects were very certain of their answers, both when they were correct and when they were incorrect. The respective mean ratings were 6.43 ($SD = 0.22$) and 6.24 ($SD = 0.19$) out of 7. In fact, as is obvious in Figure 1a, no rating for any item by any high-certainty subject was lower than 6. The mean ratings of 5.09 ($SD = 0.82$) and 4 43 (0.78) for correct and incorrect items respectively in the one-reading condition were lower than the corresponding means in the high-certainty condition, smaller $t(38) = 7.06$, $p < .001$. Although the confidence ratings in the one-reading condition averaged on the high end of the scale, they spanned its entire range, both for correct and incorrect responses (see Figure 1b). In both conditions, inspection of the distributions of ratings suggested slightly greater confidence in correct than in incorrect answers. In fact, the mean confidence ratings for correctly answered items were significantly greater than the mean ratings for incorrect items in both conditions, smaller $t(19) = 3.03$, $p < .01$.

One easily replicated finding in the error detection literature is a relationship between reading ability and error detection (Baker & Brown, 1984) Good readers are more likely than poor readers to notice when text contains anomalies and inconsistencies. A parallel relationship was not obtained here. Correlations between verbal SAT performance and awareness of when main idea questions had been answered correctly versus incorrectly (defined as the difference between each subject's confidence ratings for correct versus incorrect items) were low in both conditions, larger $|r| = 20$, $p > 20$ Moreover, there were only nonsignificant correlations between verbal ability and confidence ratings for correct items, larger $|r| = .29$, $p > .20$, and between verbal ability and confidence ratings for incorrect items, larger $|r| = 15$, $p > 50$ These failures to find significant correlations between general ability and awareness of performance are consistent with corresponding failures in Pressley et al. (in press)

SUMMARY AND INTERPRETATION

Not getting the main idea of a passage is bad enough For a reader not to know that he or she did not get it is even worse. These two deficiencies sum to overconfidence, with this occurring for the full range of reading abilities in the university sample studied here. Asking people to be really sure they answered a main idea question had two negative effects. It slowed responding to the main idea question (i.e., it increased effort expended) and increased confidence in incorrect interpretations of the main point. This pattern is similar to one we obtained previously on a learning in which college students studied two sets of sentences that differed in memorabil

(a) High-Certainty Condition
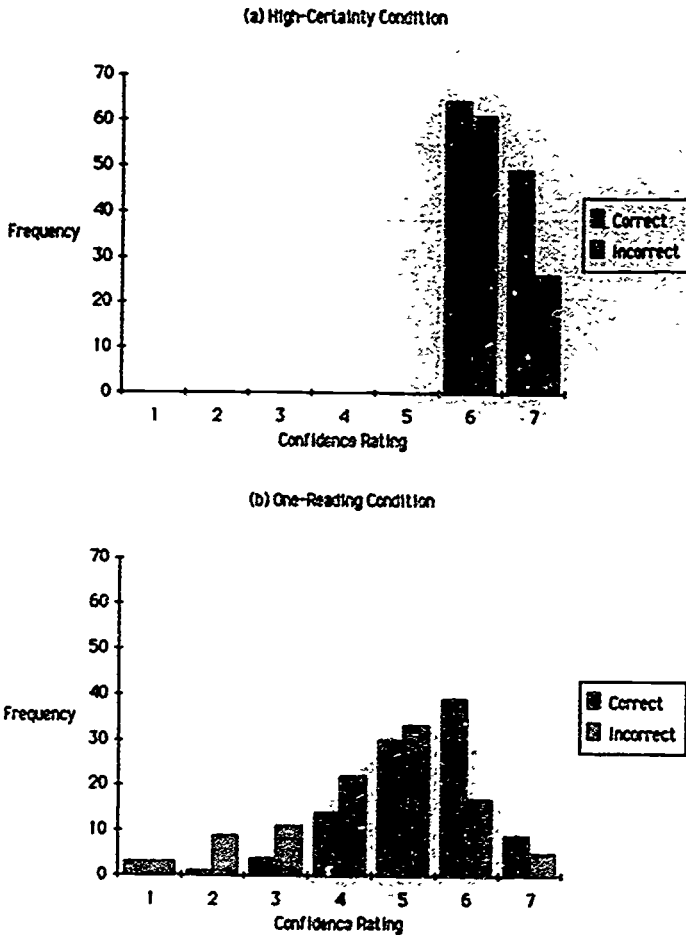

(b) One-Reading Condition

Figure 1. Distributions of certainty ratings as a function of correctness of response for the (a) high-certainty and (b) one-reading conditions.

ity (Hunter-Blanks, Ghatala, Levin, & Pressley, 1988). Subjects detected the memorability difference between sentence sets during study, reported expending more effort on the difficult sentences than on the easy sentences and predicted they would have equally good or better recall of the difficult compared to the easy sentences. In fact, however, on a subsequent test over the sentences, students recalled many more of the easy than the difficult sentences.

In both the Hunter-Blanks et al. (1988) and the present study, it is as if subjects can monitor the effort expended on the task but not the memorial or comprehension consequences of that effort. Students in the Hunter-Blanks et al. study did lower their assessment of how well they had learned the difficult sentences after they experienced a test over the sentences, however. In contrast, subjects in the present study maintained high confidence that they had comprehended the main idea of the passages even after answering test-like questions

How can this latter finding be explained? Most of the incorrect answers contained elements of the passage, embellished with additional meaning by the readers. For instance, the following were incorrect statements of the primary purpose of the example passage, all of these statements rated as likely to be correct by students who provided them (i.e., a confidence rating of 6 or 7 was provided for each one): "The historical analysis of how the architecture in Tokyo was established." "Point out the power of the press." "Say that quality can stand the test of time." "Why Wright was given so much attention for the building he built." The readers providing these responses constructed interpretations of the text, ones capturing some theme in the passage, although not the most important one. Thus, the example passage was revealing about how some of the architecture in Tokyo came about, it attested to the power of the press in enhancing a person's reputation, it told how Wright sent forth the message that his workmanship would stand challenging tests. and it specified how Wright gained a lot of attention for a building he designed. Nor are these responses atypical of incorrect answers, most of the incorrect responses consisted of themes developed in the passages. One hypothesis suggested by these incorrect answers is that so long as a reader can construct an answer to a main idea question, one that can be defended in light of some of the passage content, he or she is at risk for believing the interpretation maps well on to the main message in text.

In presenting this hypothesis, we recognize that the texts used in this investigation and in our previous studies may be special cases. In all of our work on this problem to date, the passages have been challenging and inconsiderate (Armbruster, 1984) Of course, the main ideas of more considerate texts should be more obvious and thus, more likely to be comprehended. So might the main ideas of texts that are of greater interest to readers or more consistent with their expertise. The more critical question, however, from the perspective of this investigation, is whether readers are aware of when they miss the point of such texts. A high priority should be to determine whether the comprehension monitoring problem reported here is a more general one

Even if it is not, however, the deficiency documented here is probably important, for readers are often confronted with inconsiderate texts covering content not related much to what they already know (Armbruster, 1984). Students often are required to extract the main points from difficult texts. If the only instruction provided to them is to keep working until they are very sure the main idea has been identified, that bit of instruction may do more harm than good by slowing reading, yet increasing confidence in incorrect responses. Criterion shifting alone is probably not enough to improve comprehension monitoring of inconsiderate texts.

REFERENCES

Armbruster, B B (1984) The problem of inconsiderate text In G G Duffy, L R Roehler, & J Mason (Eds) . *Comprehension instruction Perspectives and suggestions* (pp. 202–217) New York Longman.

Baker. L (1985) How do we know when we don't understand? Standards for evaluating text comprehension In D L Forrest-Pressley, G E MacKinnon, & T. G Waller (Eds), *Metacognition. cognition. and human performance,* Vol 1. *Theoretical perspectives* (pp 155–205) Orlando. FL Academic Press.

Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson, M. Kamil, R.
    Barr, & P. Mosenthal (Eds), *Handbook of reading research* (pp. 353–394). New York: Longman.
College Entrance Examination Board (1988). *10 SAT Exams* (3rd ed.). Princeton, NJ: College Entrance
    Examination Board.
Elliott-Faust, D. J., & Pressley, M. (1986). How to teach comparison processing to increase children's
    short- and long-term listening comprehension monitoring. *Journal of Educational Psychology, 78,*
    27–33.
Hunter-Blanks, P., Ghatala, E. S., Levin, J. R., & Pressley, M. (1988). Comparison of monitoring during
    study and testing on a sentence learning task. *Journal of Educational Psychology, 80,* 279–283.
Kirk, R. E (1982). *Experimental design for the behavioral sciences* (2nd ed.). Monterey, CA: Brooks/
    Cole.
Markman, E M. (1977). Realizing that you don't understand. A preliminary investigation. *Child Develop-
    ment, 48,* 986–992.
Markman, E M. (1979). Realizing that you don't understand. Elementary school children's awareness of
    inconsistencies. *Child Development, 50,* 643–655.
Markman, E M , & Gorin, L (1981). Children's ability to adjust their standards for evaluating comprehen-
    sion. *Journal of Educational Psychology, 73,* 320–325.
Pressley, M., Ghatala, E. S., Woloshyn, V., & Pinie, J (in press). Sometimes adults miss the main ideas
    and do not realize it Confidence in responses to short-answer and multiple-choice comprehension
    questions. *Reading Research Quarterly.*
Winograd, P , & Johnston, P (1982) Comprehension monitoring and the error detection paradigm. *Journal
    of Reading Behavior, 14,* 61–76.