

The relationship between comprehension and metacomprehension ability

RUTH H. MAKI, DAWN JONAS, and MARY KALLOD
North Dakota State University, Fargo, North Dakota

We investigated the relationship between the ability to comprehend text and the ability to predict future performance and to assess past performance on text. Subjects were poor at predicting performance, which may be why prediction accuracy did not relate to measures of comprehension ability. Measures of comprehension ability did relate to the accuracy with which subjects assessed their performance on tests. Better and faster comprehenders judged their relative levels of test performance over sections of text more accurately than did poorer and slower comprehenders.

Theories of comprehension frequently involve a monitoring process in which readers assess how well they are comprehending and then make self-corrections while reading (e.g., Ruddell & Speaker, 1985). If such self-corrections are critical for effective reading, the ability to make accurate assessments of learning should be related to the ability to comprehend. In addition, better comprehenders should be better at assessing how well they have performed on a test over text material. Evidence that comprehension ability is related to metacomprehension has been mixed.

Good readers can generally detect errors in text better than poorer readers can (e.g., Baker, 1985; Otero & Kintsch, 1992). However, better comprehenders predict future performance better than do poorer comprehenders in some studies (Glover, 1989; Maki & Berry, 1984) but not in others (Pressley, Snyder, Levin, Murray, & Ghatala, 1987). Overall, there is fairly good evidence that good comprehenders detect errors better than do poorer comprehenders, but the evidence with respect to predictions of performance is inconsistent.

With respect to confidence judgment accuracy, some studies have shown that higher achieving students assessed their performance more accurately than lower achieving students (Shaughnessy, 1979; Zechmeister, Rusch, & Markell, 1986). Other studies (Pressley & Ghatala, 1988; Pressley, Ghatala, Woloshyn, & Pirie, 1990) have produced nonsignificant correlations between the accuracy of confidence judgments and comprehension ability. However, these correlations were about .30, suggesting that a weak relationship between confidence judgment accuracy and verbal ability may exist.

The present study was designed to investigate the relationship between several measures of comprehension ability and both prediction and posttest confidence-judgment accuracy. One set of measures was the Multi-Media Com-

prehension Battery (MMCB; Gernsbacher & Varner, 1988). This battery involves answering questions about written, auditory, and pictorial stories. The MMCB was selected because Gernsbacher, Varner, and Faust (1990) showed that good comprehenders, as measured by the MMCB, have better access to recently comprehended information than do poorer comprehenders. Several researchers have argued that predictions of future performance are made by assessing the amount of information that can be retrieved from memory (Koriat, 1993; Morris, 1990). Therefore, better comprehenders on the MMCB should be better at predicting performance than should poorer comprehenders. Amount of retrievable information may also be important in judging the accuracy of test answers. If so, then better comprehenders should also be better at posttest confidence judgments than should poorer comprehenders. In addition to the MMCB, a more standard test of comprehension, the Nelson-Denny Reading Test (Brown, Nelson, & Denny, 1973), was used to assess comprehension ability.

We correlated each measure of comprehension ability with the accuracy of the subjects' predictions and also with the accuracy of their posttest confidence judgments. We also conducted multiple regression analyses in which several measures of comprehension ability were combined in an attempt to explain posttest confidence-judgment accuracy.

METHOD

Materials

MMCB. The MMCB (Gernsbacher & Varner, 1988) consists of three parts: written comprehension, auditory comprehension, and pictorial comprehension. Each part includes two narratives that were adapted from children's stories. The written stories are 636 and 585 words long; the auditory stories are 958 and 901 words long. The pictorial stories consist of 31 and 32 pictures. A set of 12 short-answer questions is used to test memory for each story.

Nelson-Denny Reading Test. The comprehension and reading rate portions of the Nelson-Denny Reading Test, Form C, were used. This test consists of 8 expository texts and 36 five-alternative multiple-choice questions. Comprehension is scored by giving 2 points for each correct multiple-choice answer. Reading rate is assessed by having subjects indicate how much they have read after 1 min.

We are grateful to Morton A. Gernsbacher for providing the slides and other materials for the Multi-Media Comprehension Battery. We also thank Joel Bartholomay, Nia Olson, and Jodie Rheault for testing some of the subjects. Requests for reprints can be sent to Ruth H. Maki, Department of Psychology, North Dakota State University, Fargo, ND 58105-5075.

Metacomprehension texts. Two texts taken from *Science News* were used for the metacomprehension portion of the study. One text was entitled "The Maternal Cocaine Connection" (Fackelmann, 1991). The other text concerned steroid addiction and was entitled "Pumped Up and Strung Out" (Bower, 1991). The cocaine text was 1,615 words in length, and the steroid text was 1,754 words long. Each text was broken into 12 sections with 4 to 14 sentences each. Four four-alternative multiple-choice questions were written to cover the material in each of the 12 sections.

Procedure

All subjects participated in two experimental sessions lasting about 1 h, 15 min each. The subjects completed the MNCB in one session and the Nelson-Denny test followed by the metacomprehension task in the other session. In the MNCB session, six stories were presented in the same order to all subjects, following the procedures used by Gernsbacher et al. (1990). The written stories and the test questions for all stories were presented individually to the subjects on computer monitors in small rooms. The auditory and pictorial stories were presented in a larger room to groups of 2 to 5 subjects. Immediately following each story, each subject saw 12 questions on a computer screen and was allowed 20 sec to write an answer to each question on a numbered answer sheet.

The Nelson-Denny/metacomprehension session began with the timed Nelson-Denny test, which was administered in groups varying in size from 2 to 5 subjects. The standardized test instructions were used. The subjects were given the test booklet and told to begin reading the first passage. One minute after they began to read, the subjects were asked to indicate the line that they were reading. They then continued reading passages and answering questions for another 19 min.

Instructions were given for the metacomprehension portion of the experiment. The subjects were randomly assigned to read either the cocaine or the steroid text. Each text was presented on a computer monitor, one sentence at a time. Presentation was self-paced, and the subjects pressed the space bar to move on to the next sentence when they had read each sentence. At the end of a section of text, roughly corresponding to a paragraph, a prediction rating scale was presented. This was accompanied by the question, "How poorly will you do on the section of text that you just read?" A 6-alternative rating scale was used, with 6 = "Very poorly" and 1 = "Very well." The question was framed in terms of "poorly" rather than "well" because Maki & Serra (1992) found that "poorly" produced higher confidence-judgment accuracy than did "well." After all 12 sections were read and rated, the subjects saw the four multiple-choice questions for each section. Questions for the sections were presented in the same order in which the sections had been read, but questions within a section were presented in random order. After the four questions from a section of text had been answered, a confidence-judgment rating scale was presented. It was accompanied by the question "How poorly did you do on the set of questions that you just answered?" with 6 = "Very poorly" and 1 = "Very well." Questions from the next section were then presented and answered, followed by a confidence judgment for them. The computer was programmed to present the text, ratings, and multiple-choice questions by using the Micro Experimental Laboratory (MEL; Schneider, 1988).

Subjects. A total of 75 volunteers from introductory psychology classes at North Dakota State University participated in the study. Forty-two subjects participated in the MNCB portion first, and 33 participated in the metacomprehension portion first. Thirty-eight of the subjects read the cocaine text, and 37 read the steroid text.

RESULTS

Descriptive Statistics

Comprehension measures. The MNCB yielded three scores for each subject: a score on the written stories, a score on the auditory stories, and a score on the pictorial stories. Each of the 12 questions for each story was given a score from 0 to 3, using the scoring manual developed by Gernsbacher and Varner (1988). Two independent raters scored a randomly selected 18 of the recall proto-

cols. For the written stories, the Pearson r correlation between the two raters' scores was .991; for the auditory stories, the r was .986; and for the pictorial stories, the r was .965. Mean scores for our entire sample on the written, auditory, and pictorial stories were 19.86 ($SE = .58$), 18.54 ($SE = .65$), and 16.84 ($SE = .47$), respectively, out of 36. Our mean on the written test did not differ from Gernsbacher et al.'s (1990) mean of 20.75 [$t(74) = -1.53, p > .05$ (the significance level used for all analyses)], but our means on the auditory and pictorial tests were significantly lower than Gernsbacher et al.'s means of 20.97 and 20.48 [$ts(74) \leq -3.84$]. Mean performance on the Nelson-Denny comprehension measure was 38.64 ($SE = 1.04$), placing our subjects at the 45th percentile. The mean proportion correct on the *Science News* texts was .685 ($SE = .015$).

Two measures of reading rates were also taken. The mean reading rate on the Nelson-Denny was 205.3 ($SE = 5.62$), placing our subjects at the 26th percentile. The second reading rate measure was taken while the subjects read the *Science News* texts. The mean number of milliseconds per word was 149.1 ($SE = 4.12$), or 402 words per minute.

Metacomprehension measures. Ratings made by the subjects to predict performance were reversed so that high ratings went with predictions of high performance. This reversed mean prediction rating on the 6-point scale was 4.03 ($SE = .088$), with 6 indicating that subjects thought they would perform very well. Our main measure of metacomprehension accuracy was the gamma correlation (Goodman & Kruskal, 1954), which treats the ratings as ordinal data (a practice recommended by Nelson, 1984) and indicates the relationship between the ratings and performance. For each subject, two contingency tables were created. One contained the prediction ratings for each section of text (ranging from 1 to 6) and the total scores on the four questions for each section (ranging from 0 to 4); the other contained the confidence judgments and scores. We determined whether there was a relationship between each set of ratings and the scores with gamma, which varies from -1.0 (indicating a perfect negative relation) to $+1.0$ (indicating a perfect positive relation). Although gammas can be calculated from ordinal data, the gammas themselves can be treated with parametric statistics. If there is no relationship between two variables, gammas are normally distributed around zero.

Prediction gammas could be determined for 73 of the 75 subjects; the other 2 subjects gave the same prediction rating to each section of text. The mean prediction gamma in the present experiment was .114 ($SE = .050$). A single-sample t test indicated that this gamma was significantly greater than that expected by chance [$t(72) = 2.28$].

The mean confidence judgment ratings given after the test questions had been answered was 4.24 ($SE = .080$) on the 6-point scale. Gammas for confidence judgment accuracy could be computed for 73 of the 75 subjects. Again, 2 subjects were dropped from this analysis because they gave the same confidence judgment rating to all text

Table 1
Intercorrelations (Pearson r) Among Measures of Comprehension

	MMCB Auditory	MMCB Pictorial	<i>Science News</i> Score	Nelson- Denny Comprehension	Nelson- Denny Rate	Rate on <i>Science News</i> Texts
MMCB written	.626*	.164	.538*	.455*	.334*	-.106
MMCB auditory		.274	.457*	.417*	.134	.001
MMCB pictorial			.227	.146	.021	-.132
<i>Science News</i> score				.507*	.139	.118
Nelson-Denny comprehension					.362*	-.150
Nelson-Denny rate						-.214†

* $p < .05$. † $p < .10$.

sections. In contrast to the low accuracy of prediction ratings, the accuracy of these confidence judgments was quite high [$\gamma = .551$ ($SE = .043$)]. A single-sample t test indicated that this value was significantly greater than zero [$t(72) = 12.85$, $p < .01$].

Correlations

Comprehension measures. Table 1 shows the Pearson r correlations among the measures of comprehension. Generally, the measures of verbal comprehension were correlated, including the written and auditory portions of the MMCB, the comprehension scores on the Nelson-Denny Reading Test, and the scores on the *Science News* texts. However, the comprehension of pictorial stories from the MMCB was not correlated with the other comprehension measures. The two measures of reading rate (Nelson-Denny reading rate and the milliseconds per word on the *Science News* texts) tended to be negatively correlated because a high score on the Nelson-Denny meant fast reading and a low score on the text reading rate meant fast reading. The Nelson-Denny reading rate measure correlated with the measures of written comprehension, but the reading rate measure from the metacomprehension texts did not correlate with the other measures of comprehension.

Comprehension ability and metacomprehension. Table 2 shows the Pearson r correlations among the various measures of comprehension ability and the prediction and posttest confidence-judgment gammas. The only measure of comprehension ability that correlated with prediction accuracy was the auditory score on the MMCB, although the Nelson-Denny reading rate score also showed some tendency to correlate positively with prediction accuracy. In contrast, almost all of the measures of comprehension ability correlated with the confidence-judgment gammas. Generally, better comprehenders and faster readers gave confidence judgments that related more closely to actual performance than did poorer comprehenders and slower readers.

Reliabilities. One reason why the ability to predict performance may not have correlated with comprehension ability is because the measure of prediction ability was not reliable; the measure of confidence-judgment accuracy may have been more reliable. Internal reliabilities were assessed with Cronbach's alpha. For the prediction ratings, alpha was .917; for the confidence-judgment ratings, it was .887; and for the test scores, alpha was .741. A

similar measure was computed for the gammas in order to see whether there was internal consistency in our measurement of the subjects' abilities to predict future and to assess past test performance. A random set of six text paragraphs was selected, and separate prediction and confidence-judgment gammas were calculated on the basis of them. This procedure was repeated 12 times so that each subject had 12 different prediction and confidence-judgment gammas. For predictions, the range of these 12 mean gammas was $-.040$ to $+.339$, with Cronbach's alpha equal to .904. For confidence judgments, the range was $+.471$ to $+.625$, with an alpha of .920. Thus, both the prediction and confidence-judgment gammas showed high internal consistency.

Multiple Regression Analyses

Prediction accuracy. An original goal of this study was to explain the variance in the subjects' abilities to predict their performance on the *Science News* texts with various measures of comprehension ability. However, the subjects were very poor at predicting performance, and the variance in their prediction gammas may have been mostly error variance. Thus, further analyses using prediction accuracy are not reported.

Confidence judgment accuracy. Multiple regression analyses were conducted to explain the variance in the accuracy of the subjects' posttest confidence judgments. The predicted variable was each subject's gamma correlation, which was our measure of confidence-judgment accuracy. The first set of predictor variables that was en-

Table 2
Correlations (Pearson r) Between
Metacomprehension Accuracy and Comprehension Ability

Comprehension Measure	Type of Metacomprehension Accuracy	
	Prediction	Confidence- Judgment
<i>Science News</i> texts		
Score	.046	.332*
Reading rate	.027	-.234*
MMCB		
Written	.084	.433*
Auditory	.234*	.346*
Pictorial	.020	.026
Nelson-Denny		
Comprehension	.013	.307*
Reading rate	.204†	.326*

* $p < .05$. † $p < .10$.

Table 3
Multiple Regression Analysis for the
Accuracy of Confidence Judgments

Variable Set Added	Model <i>df</i>	Model <i>R</i> ²	Increment in <i>R</i> ²
Verbal comprehension	4	.2190*	.2190*
Pictorial comprehension	5	.2264*	.0074
Reading rate	7	.2998*	.0734*

**p* < .05.

tered into the regression model was the verbal comprehension variable set, including the written and auditory scores on the MMCB, the comprehension score from the Nelson-Denny Reading Test, and the score on the questions from the *Science News* texts. As shown in Table 3, these variables explained a significant amount of variance. The relatively independent MMCB pictorial comprehension score did not add a significant increment in the amount of variance explained. The final set of predictor variables consisted of the two reading rate measures: the Nelson-Denny rate and milliseconds per word for the *Science News* texts. These two time measures increased the amount of variance explained. The seven-variable model explained almost 30% of the variance in confidence-judgment accuracy for a multiple *R* of .547.

DISCUSSION

The subjects were very poor at predicting their performance. This is consistent with other studies using these texts. Under conditions similar to the present ones, Maki, Foley, Kajer, Thompson, and Willert (1990) found a mean prediction gamma of .176. Because of the low prediction accuracy with these texts, the variance in the prediction gammas may have been mostly noise. If so, these gammas would not correlate with other measures.

In contrast, posttest confidence judgments were accurate and were related to the measures of verbal comprehension ability and to speed of comprehension. Better comprehenders and faster readers made more accurate judgments about their prior test performance. The finding that posttest confidence-judgment accuracy was related to the measures of verbal comprehension ability is consistent with some past research (Shaughnessy, 1979; Zechmeister et al., 1986). However, the present conclusions are different from those of Pressley's studies (Pressley & Ghatala, 1988; Pressley et al., 1990), in which no relationship between posttest confidence judgments and verbal ability was found. Still, our correlations are similar in magnitude to those found in Pressley's studies, but more degrees of freedom make our correlations significant, whereas their correlations were not significant.

The present results fit well with Koriat's (1993) recent theory of feeling-of-knowing judgments. He argues that subjects base recognizability on how much partial information can be recalled. Subjects who are more able to access this partial information are better at predicting whether they will recognize answers. In the present case, subjects who comprehend better (and presumably who have better access to stored information) could judge whether they had recognized correct answers better than could subjects who comprehend less well. This finding also fits with Gernsbacher et al.'s (1990) theory of comprehension in that good comprehenders have more access to stored information than do poor comprehenders. These theories about the relationship of verbal abilities to metacomprehension would predict a relationship of verbal ability both to prediction ability and to the ability to make accurate confidence judgments. However, because the prediction task was too difficult for our subjects, we could not expect to find any relationship. In contrast, we did show that better comprehenders are better at assessing their levels of prior performance.

REFERENCES

- BAKER, L. (1985). Differences in the standards used by college students to evaluate their comprehension of expository prose. *Reading Research Quarterly*, **20**, 297-313.
- BOWER, B. (1991). Pumped up and strung out. *Science News*, **140**, 30-31.
- BROWN, J. L., NELSON, M. J., & DENNY, E. C. (1973). *The Nelson-Denny Reading Test*. Chicago: Riverside.
- FACKELMANN, K. A. (1991). The maternal cocaine connection. *Science News*, **140**, 152-153.
- GERNSBACHER, M. A., & VARNER, K. R. (1988). *The multi-media comprehension battery* (Tech. Rep. No. 88-07). Eugene, OR: Institute of Cognitive and Decision Sciences.
- GERNSBACHER, M. A., VARNER, K. R., & FAUST, M. E. (1990). Investigating differences in general comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 430-445.
- GLOVER, J. A. (1989). Improving readers' estimates of learning from text: The role of inserted questions. *Reading Research & Instruction*, **28**, 68-75.
- GOODMAN, L. A., & KRUSKAL, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, **49**, 732-764.
- KORIAT, A. (1993). *How do we know that we know? The accessibility model of the feeling of knowing* (IPDM Report No. 104). Haifa, Israel: The Institute of Information Processing and Decision Making.
- MAKI, R. H., & BERRY, S. A. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 663-679.
- MAKI, R. H., FOLEY, J. M., KAJER, W. J., THOMPSON, R. C., & WILLERT, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 609-616.
- MAKI, R. H., & SERRA, M. (1992). The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 116-126.
- MORRIS, C. C. (1990). Retrieval processes underlying confidence in comprehension judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 223-232.
- NELSON, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, **95**, 109-133.
- OTERO, J., & KINTSCH, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science*, **3**, 229-235.
- PRESSLEY, M., & GHATALA, E. S. (1988). Delusions about performance on multiple-choice comprehension tests. *Reading Research Quarterly*, **23**, 454-464.
- PRESSLEY, M., GHATALA, E. S., WOLOSZYN, V., & PIRIE, J. (1990). 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*, **25**, 232-249.
- PRESSLEY, M., SNYDER, B. L., LEVIN, J. R., MURRAY, H. G., & GHATALA, E. S. (1987). Perceived readiness for examination performance (PREP) produced by initial reading of text and text containing adjunct questions. *Reading Research Quarterly*, **22**, 219-236.
- RUDDELL, R. B., & SPEAKER, R. B., JR. (1985). The interactive reading process: A model. In H. Singer and R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (3rd ed., pp. 751-793). Newark, Del.: International Reading Association.
- SCHNEIDER, W. (1988). *Micro Experimental Laboratory (MEL)* [Computer program]. Pittsburgh: Psychology Software Tools.
- SHAUGHNESSY, J. J. (1979). Confidence-judgment accuracy as a predictor of test performance. *Journal of Research in Personality*, **13**, 505-514.
- ZECHMEISTER, E. B., RUSCH, K. M., & MARKELL, K. A. (1986). Training college students to assess accurately what they know and don't know. *Human Learning*, **5**, 3-19.

(Manuscript received June 17, 1993;
 revision accepted for publication August 13, 1993.)