Individual Differences in
Comprehension Monitoring Ability during Reading

Christopher A. Kurby (ckurby@mail.psyc.memphis.edu)
Yasuhiro Ozuru (y.ozuru@mail.psyc.memphis.edu)
Danielle S. McNamara (d.mcnamara@ mail.psyc.memphis.edu)

Psychology Department, University of Memphis
Memphis, TN 38152 USA

Abstract

The goal of this experiment was to investigate readers' ability to monitor their processing difficulty during reading and individual differences in this ability. Participants read a text one sentence at a time and made judgments of learning, judgments of sentence difficulty, or made no judgment. Participants' metacognitive judgments were correlated with their processing difficulty only for high skilled readers. In addition, attending to different metacognitive judgments did not increase accuracy on comprehension questions compared to a read only group. The results suggest that comprehension monitoring is a skilled process, but may not be sufficient for skilled comprehension.

Keywords: metacognition; comprehension; individual differences; comprehension monitoring.

Introduction

Students' metacomprehension abilities have drawn the attention of reading researchers and educators, at least partially because of the clear implications of these abilities on learning. That is, the degree to which students are aware of their comprehension difficulties during reading should determine how effectively they can adjust their study techniques accordingly (e.g., Thiede, Anderson, & Therriault, 2003).

This paper explores two specific questions related to metacomprehension. The first question regards the status of the metacomprehension process itself. Can readers accurately monitor their understanding, and if so, what kind of information do readers use when monitoring their comprehension? The second question regards whether instructions to report either judgments of learning or judgments of processing difficulty influence comprehension of written materials. That is, does requiring the reader to focus on metacomprehension processes improve comprehension?

With regard to the first question, research to date has revealed rather complex and inconclusive pictures of students' ability to metacognitively monitor comprehension. That is, readers' ability to monitor comprehension appears to be rather poor or at least fair depending on the research technique used to assess the ability.

For instance, Pressley, Ghatala, Wolosyn, and Pirie (1990) showed that readers often do not look back at the text when answering comprehension questions and as a result they tend to provide incorrect answers, indicating their tendency to overestimate their comprehension.

Similarly, studies using a contradiction paradigm have shown that readers are often not aware of the presence of the contradiction when explicitly probed about the contradiction (Epstein, Glenberg, & Bradley, 1984). These findings are rather surprising given that text comprehension research indicates that readers adjust reading speed accordingly when they encounter difficulty and/or contradictions (Haberlandt & Graesser, 1985). Thus, there may be a dissociation between on-line processing behavior and awareness (meta-cognition) of readers' own processing behavior (i.e., reading time).

On the other hand, more recent research on metacomprehension using a judgment of learning paradigm has indicated that some readers under some circumstances successfully monitor their comprehension. When absolute judgments were made, at least relatively skilled readers have a moderate ability to estimate how well they will do on a test based on their understanding of previous text material (Maki, Shields, Wheeler, & Zacchilli, 2005). That is, when readers' judgments of learning are assessed in terms of the degree to which they can accurately predict question answering performance based on a given text, skilled readers show relatively high levels of judgment accuracy. This line of research suggests that readers' ability to monitor their comprehension may depend on individual differences, such as reading skill.

The notion that some readers are better than other readers in terms of metacomprehension leads to the question of how "skilled" readers monitor their comprehension of the text. A study by Rawson and Dunlosky (2002) offers an explanation. They showed, using a judgment of learning paradigm, that readers' judgments of learning are higher in magnitude for texts that have high cohesiveness than texts with low cohesiveness. The more cohesive the text (i.e., easier to understand), the more confident they were of success on a following test of the text content. This finding led them to conclude that students' assessment of comprehension is to some extent based on ease of processing, operationalized as cohesiveness of text. However, it is not entirely clear

whether their participants' judgments were directly based on analytical observations of the stimulus texts (cohesiveness) or based on changes in processing difficulty caused by changes in cohesiveness. Interestingly, Rawson and Dunlosky (2002) argued that judgments of learning are not based on reading time per se. This conclusion was based on a lack of correlation between judgments of learning and reading time. However, this finding is somewhat incongruous because their data also showed that reading time of the texts was longer for low cohesion texts than for high cohesion texts, which is consistent with text processing literature that shows readers often slow down when they encounter difficulty (e.g., Haberlandt & Graesser, 1985). In any event, they did not fully explore the relations among reading time, processing difficulty (text difficulty), and judgment ratings. Thus, gaining insight into readers' ability to monitor comprehension requires an analysis of the relationship among stimulus features (text difficulty), on-line processing (e.g., eye-tracking, reading time), and comprehension monitoring.

Hence, the goal of this experiment was to investigate the extent to which readers can monitor their processing difficulty during comprehension. Given the findings reported earlier (e.g., Maki et al., 2005; Rawson & Dunlosky, 2002), it is plausible that readers, to some extent, are able to accurately evaluate their online comprehension processes (reading time), not just text difficulty, even though there may be individual differences in this ability (e.g., Maki et al., 2005). That is, can participants use information regarding their processing fluency (i.e., reading time, text difficulty) when making metacognitive evaluations of their understanding? In the current experiment, we investigated this possibility with respect to reading ability. Thus, we explored individual differences (i.e., reading skill) in readers' ability to monitor their comprehension based on both objective text difficulty (i.e., reading ease) and processing difficulty as experienced during reading (i.e., reading time).

Given that skilled readers can somewhat successfully monitor comprehension, a question remains with respect to how comprehension monitoring contributes to better comprehension. On the one hand, comprehension monitoring may benefit comprehension because this would lead to efficient remedial processing in case of encountering difficulty. According to this view, comprehension monitoring per se is not sufficient to increase comprehension unless readers engage in an appropriate remedy or *fix-up strategies* (McNamara, in press). An alternative view is the possibility that the act of monitoring alone facilitates comprehension. That is, when monitoring comprehension, a reader engages in an active evaluation of their understanding, which may promote deeper processing of the text material (Graesser, Singer, & Trabasso, 1994). In this paper, in addition to individual differences in comprehension monitoring, we explored whether explicitly directing readers to attend to different types of comprehension monitoring influences comprehension of the text content. Thus, in the current study, participants read a text under one of three different sets of instructions. One group was asked to report a judgment of learning after each sentence (i.e., JOL), a second group was asked to report judgments of how difficult it was to understand each sentence (i.e., JOD), and one group simply read for comprehension (i.e., read only). All groups then answered multiple-choice questions about the text. If participants are able to use information regarding their processing fluency to to monitor their comprehension difficulty, then JOLs and JODs should be correlated with reading time for each sentence. If engaging in monitoring tasks facilitates comprehension then accuracy on the comprehension questions will be greater for the two judgment groups compared to the read only group.

Method

Participants and Design

Participants were 32 undergraduate students enrolled in introductory psychology courses at the University of Memphis. They received course credit for participating in the experiment. Participants were randomly assigned to one of the three reading conditions (i.e., read only, judgment of learning, or sentence difficulty judgment). The sentence difficulty judgment condition had 10 participant and the other two conditions had 11 participants each.

Materials

The text was an expository passage entitled "The Lower Level Brain Structures" which described the structure and functions of the lower brain structure. The text was obtained from a chapter on the brain in a psychology text book "Psychology, Myers in Modules, sixth edition" (Myers, 2001). The text was modified to increase variability in the processing difficulty of each sentence by eliminating/re-ordering sentences, replacing words, and/or modifying the sentence structure. The final text was comprised of 62 sentences with sentence lengths varying from 4 to 49 words per sentence. The Flesch Reading Ease for each sentence was calculated using Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004). The scores for the 62 sentences varied from 0 to 96.7. The frequency distribution of sentences with different reading ease levels is presented in Table 1.

Table 1. Frequency distribution of sentences as a function of reading ease level

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

The total number of words in the text was 1188. Fifty-nine four-option multiple choice comprehension questions were constructed based on the text. The questions included 39 text-based questions (i.e., questions based on a single sentence), 14 bridging inference questions (i.e., questions based on multiple sentences), and 6 vocabulary questions. Care was taken to ensure that these questions probed for a variety of information contained in the different sections of the text. Participants' reading skill was assessed using the Gates-MacGinitie Reading Test for grade level 10/12 form S (GMRT).

Procedure

Participants were first administered the Gates MacGinitie reading ability test. They were given a 20 minute time limit. After taking the reading ability test, participants read the text one sentence at a time presented by the E-prime (2000) computer program on a notebook computer.

Prior to reading the text, participants received different instructions depending on the condition. The *Read Only* condition participants were instructed to read the text carefully one sentence at a time. The judgment of learning condition (*JOL*) participants were asked to indicate how well they would answer a question based on each sentence using a 4-point scale (1=likely to be wrong, 4=likely to be correct). The sentence difficulty judgment condition (*JOD*) participants were asked to indicate their subjective difficulty estimate of each sentence using a 4-point scale (1=very easy, 4=very difficult) immediately after reading each sentence. Reading time of each sentence and the judgment responses for the judgment conditions were recorded.

In all the three conditions, participants were informed that they would answer comprehension questions based on the text after they were finished reading. Immediately after reading the text and performing the judgment task, participants answered the multiple choice questions in the absence of the text. The questions were presented one at a time on a notebook computer using the E Prime (2000) program. They indicated their answer by pressing an appropriate key corresponding to one of the answer options. Question answering was self-paced.

Results and Discussion

There are two subsections to the results. The first section reports the analysis of the relationship between judgments and reading time, and the effects of individual differences on the information that participants attend to during meta-comprehension. The second section reports the analysis of the effect of the judgment tasks on text comprehension. In all the analyses reported, we are interested in the effect of judgment condition (i.e., type of judgment reported) and reading skill on various dependent variables (e.g., relations between reading time and judgment ratings). Thus, we used a median split technique to create a quasi experimental variable to represent level of reading skill based on performance on the GMRT.

Individual Differences in Comprehension Monitoring

We were interested in the extent to which skilled and less skilled readers, as measured by performance on the GMRT, can monitor their cognitive processing during comprehension, rather than after comprehension (as in Rawson & Dunlosky, 2002). We used each participant's reading time per word per sentence as a measure of their processing difficulty/ease. If their metacognitive judgments are based on their subjective experience of processing difficulty then the judgments should be correlated with reading time. That is, ratings of difficulty should increase as reading time increases. JOLs, on the other hand, should decrease as reading time increases because longer reading times indicate that the expected question based on the sentence would be more difficult. For ease of exposition, the JOLs were reverse scored such that the higher the rating, the lower the prediction of performance. In this way, we predict positive correlations between reading time and ratings (both JOLs and JODs) if participants use information about their processing fluency to make metacognitive judgments of their comprehension difficulty.

In order to analyze the correlation between reading time and judgments for each judgment group we followed a procedure developed by Lorch and Myers (1990). In these analyses, we conducted a separate regression on each participant's judgments using reading time as the predictor and judgment as the criterion variable. The standardized beta weight was then extracted for each participant. The extracted beta weights were subjected to single-sample t-tests to determine whether or not they were significantly different from zero. The mean beta weights for the JOL group ($B = .12$) and the JOD group ($B = .19$) were both significantly greater than zero, $t(10) = 2.63$, $p < .05$ and $t(9) = 2.63$, $p < .05$ respectively. This suggests that offline JOLs and JODs are to some extent based on participants' experience of processing difficulty during reading. It appears then that participants have some ability to monitor their own understanding during comprehension.

We also investigated to what extent readers' ability to monitor processing difficulty was related to reading skill. That is, reading skill may comprise one's ability to monitor their comprehension online. As such, we replicated the analysis above but conducted separate t-tests against zero for each reading skill group (high vs. low). The beta weights are presented in Table 2. These represent the correlation between the reader's reading time and judgments. A positive value means that reading time increases (i.e., slows down) as judgments increase. The t tests revealed that only the high skilled readers showed significant correlations between reading time and judgments (JOL-Low: $t(5) < 1$; JOL-High: $t(4) = 3.85$, $p < .05$; JOD-Low: $t(3) < 1$; JOD-High: $t(5) = 2.92$, $p < .05$). This suggests that the use of information regarding processing fluency to monitor comprehension is related to reading skill. That is, whereas skilled readers' JOD and JOL estimations are somewhat congruent with their own reading difficulty (operationalized as reading time), the judgments of less-skilled readers are not congruent with their reading time, hinting at a fundamental difference in monitoring ability between skilled and less skilled readers.

Table 2. Mean Beta Weights for each Group and Skill.

Condition	Reading Skill	Beta Weight	Beta Weight *SE*
JOL	Low	.04	.05
	High	.21	.06
JOD	Low	.08	.11
	High	.27	.09

Our data suggest that readers differ in their monitoring of their online processing. Rawson and Dunlosky (2002), however, suggest that metacognitive judgments may be more directly predicted by text difficulty rather than processing speed (i.e., reading time) itself. Indeed, it is unclear here how much reading time (a cognitive measure) contributes to metacognitive judgments over and above text difficulty. To address this issue, we conducted item-based analyses using hierarchical linear regressions on the mean judgments per sentence to investigate the contribution of reading time (RT) to judgments over and above a measure of sentence difficulty (Flesch Reading Ease: FRE). Specifically, we calculated average judgments for each sentence across participants separately for high and low skilled readers in each condition (JOL and JOD). Then, using mean judgment as the criterion variable, we performed hierarchical linear regressions using average reading time of each sentence and FRE of each sentence as the predictor variables. Flesch Reading Ease (FRE) was entered in the first step and average reading time per word (RT) was entered in the second step. We were interested in the ΔR^2 for reading time for each analysis. A separate analysis was run for each of the four groups (condition x reading skill). The total R^2 and ΔR^2 for each step and

predictor (FRE, RT) for each analysis are presented in Table 3.

Table 3. Total R^2 and ΔR^2 for each step, condition, and reading skill

Condition	Reading Skill	R^2		ΔR^2
		FRE (step 1)	RT (step 2)	
JOL	Low	.000	.000	.000
	High	.230*	.232*	.002
JOD	Low	.068*	.070*	.002
	High	.159*	.245*	.086*

The first column of Table 3 (i.e., step 1) indicates that our measure of sentence difficulty was a significant predictor of judgments for all groups except for the JOL-Low skilled group. This finding replicates Rawson and Dunlosky's (2002) results in that when entered in isolation, text difficulty accounted for a significant amount of variance in participants' judgment ratings of the sentence. However, looking at the results reported in the third column (i.e., ΔR^2), we see that the skilled readers in the JOD condition based their judgments on their reading time over and above text difficulty. These readers appear to base their judgments on both text difficulty and on their assessment of their processing difficulty at the time of reading. This suggests that at least skilled readers take into consideration the subjective difficulty of their text processing (i.e., reading time) in addition to the text features when making sentence difficulty judgments. In contrast, skilled readers' JOLs were predicted by objective text difficulty alone. Finally, less skilled readers' judgments, in particular in the JOL condition, were related to neither sentence difficulty nor reading time in a systematic way.

Effect of Judgment Task on Comprehension

In this section we explored whether or not asking participants to pay attention to metacognitive activities (i.e., asking to make overt judgments) during reading increases comprehension. Table 4 presents comprehension performance (proportion correct on multiple choice comprehension questions) as a function of reading skill and conditions. A 3 (condition: read only vs. JOL vs. JOD) x 2 (reading skill: high vs. low) ANOVA was conducted with mean accuracy on the multiple choice test as the dependent measure. The only significant effect was a main effect of reading skill such that accuracy was higher for the high skilled readers ($M = .62$) than the low skilled readers ($M = .43$), $F(1, 26) = 30.19$, $MSE = .283$, $p < .001$. There was no significant effect of condition on accuracy. This shows that asking participants to pay attention to metacognitive activities during reading in itself did not increase their comprehension of the text. Instead,

text comprehension was influenced by a pre-existing individual difference, level of reading skill as measured by the GMRT.

Table 4. Mean accuracy for each condition and reading skill (SD in parentheses)

Condition	Reading Skill	
	Low	High
Read only	.50 (.15)	.63 (.09)
JOL	.36 (.09)	.63 (.09)
JOD	.43 (.06)	.60 (.07)

We conducted a follow-up analysis on the reading times to test the possibility that engaging in metacogitive evaluations of processing fluency requires additional processing time relative to normal reading. That is, reading times should be slower for the judgment groups than the read only group if attending to metacognitive assessments requires additional processing. Table 5 presents average reading times as a function of reading skill and condition.

Table 5. Mean reading time per word (ms) as a function of condition and reading skill

Condition	Reading Skill	
	Low	High
Read only	716 (360)	378 (63)
JOL	485 (188)	456 (143)
JOD	466 (213)	445 (84)

Note: SD's are in parentheses

A 2 (reading skill) x 2(condition) ANOVA indicated that no effect was statistically significant, except for a marginal effect of reading skill, $F(1, 26) = 3.10$, $p = .09$, indicating that skilled readers have some tendency to read texts faster. An important finding, however, is the null effect of condition. This essentially shows that readers did not spend extra time to perform the judgment task beyond normal reading processes. We will revisit this issue later in the discussion section.

Discussion

The goal of this study was to examine readers' ability to use information regarding their online processing fluency to monitor their comprehension. In addition, we examined whether or not asking participants to make overt metacognitive judgments during reading enhances comprehension. With respect to these goals, we investigated whether monitoring ability is moderated by reading skill and the effect of monitoring on comprehension.

We found that readers' evaluations of their comprehension correspond to their online processing fluency, but that reading skill moderated this ability. That is, only readers with high reading ability showed evidence that they can monitor their online comprehension difficulty using their reading speed as an indicator. In addition, we found evidence that replicates and extends work by Rawson and Dunlosky (2002). Readers tended to base their metacognitive judgments of their comprehension mostly on textual difficulty but also, in the JOD condition, based their judgments on an assessment of their cognitive processes. This result suggests a divergence between the more traditional judgments of learning and the judgments of comprehension difficulty that we have included in this study. That is, it appears that the tasks may tap different types of processing. The JODs appear to be made on the objective difficulty of the text and the difficulty experienced while reading. This makes sense given that the task was for the participants to make the judgments based on their own experience of difficulty. As such, difficulty judgments, at least for high skilled readers, appear to be partly based on processing fluency (i.e., reading time). The JOLs appear to be almost entirely based on the difficulty of the text itself, rather than an assessment of cognitive processing. From our data, it is unclear what aspect of the sentences participants focus on to predict their probability of success on a test question. The Flesch Reading Ease measure takes into account both word length and sentence length. It is possible that skilled readers base their judgments on the extent to which they are familiar with the longer words or on the sentence length. More work must be done to further explore the source of these judgments.

Contrary to a somewhat intuitive expectation that comprehension monitoring involves deeper processing of texts, the accuracy data suggest that comprehension did not benefit from engaging in monitoring tasks, regardless of reading skill. Yet, this conclusion may not be fully warranted because the reading time data suggest that there are no differences in reading behavior among the judgment groups. That is, if participants were engaging in metacognitive evaluations in addition to normal reading, then their reading time should have been slower in the two judgment conditions than the read only condition. However, the means in Table 5 point to a possibility that some difference in processing, albeit statistically undetected, may exist between the read only and the two judgment conditions, indicating additional processing. Table 5 shows that high skilled readers took around 70msec longer per word in the JOL and JOD conditions. Given that the average sentence length was 19.16 words, this means that high skilled readers were taking an extra 1.3 seconds per sentence, on average. This suggests that the high skilled readers were making their judgments in addition to reading the text, rather than solely reading. We acknowledge that the sample sizes are relatively small, however, given the significant effects of reading skill it is clear that we had enough power to detect differences. Nonetheless, we are currently collecting more data to ensure this is the case.

One explanation for why monitoring alone did not change the comprehension scores may be that participants were not given explicit instructions to take remedial measures when they estimated their difficulty to be high or their learning to be low. Giving participants strategies to overcome comprehension difficulty, or textual difficulty, may increase comprehension when participants' ability to monitor is high. That is, participants who recognize that they are having comprehension difficulties may be taught how to paraphrase the sentence or to connect it back to the previous discourse. Our technique of measuring participants monitoring ability could be used to inform automated reading strategy training systems such as iSTART (McNamara, Levinstein, & Boonthum, 2004). A system such as this could be implemented to teach students reading strategies that could be employed when the student feels that he/she is not understanding. It is one future goal of this work to test the extent to which the combination of metacognitive tasks and reading strategies for comprehension remediation improves understanding.

Acknowledgments

References

E-Prime 1.0 [Computer software]. (2000). Pittsburgh, PA: Psychology Software Tools.

Epstein, W., Glenberg, A.M., & Bradley, M.M. (1984). *Memory & Cognition, 12,* 355-360

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

Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371-395.

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

McNamara, D.S. (in press). *Reading Comprehension Strategies: Theories, Interventions, and Technologies*. Mahwah, NJ: Erlbaum.

McNamara, D. S., Levinstein, I. B. & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, & Computers, 36,* 222-233.

Pressley, M., Ghatala, E. S., Woloshyn, V. E., & Pirie, J. (1990). Sometimes adults miss the main idea and do not realize it: confidence in response to short answer and multiple choice comprehension questions. *Reading Research Quarterly, 25,* 232-249.

Myers, D. G. (2001). *Psychology, Meyers in Modules, sixth edition*. New York, NY: Worth Publishers.

Rawson, K.A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 69-80.

Thiede, K. W., Anderson, M.C., & Therriault (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*, 66-73.