

Accuracy of Metacognitive Monitoring Affects Learning of Texts

Keith W. Thiede, Mary C. M. Anderson, and David Therriault
University of Illinois at Chicago

Metacognitive monitoring affects regulation of study, and this affects overall learning. The authors created differences in monitoring accuracy by instructing participants to generate a list of 5 keywords that captured the essence of each text. Accuracy was greater for a group that wrote keywords after a delay (delayed-keyword group) than for a group that wrote keywords immediately after reading (immediate-keyword group) and a group that did not write keywords (no-keyword group). The superior monitoring accuracy produced more effective regulation of study. Differences in monitoring accuracy and regulation of study, in turn, produced greater overall test performance (reading comprehension) for the delayed-keyword group versus the other groups. The results are framed in the context of a discrepancy-reduction model of self-regulated study.

Many models of self-regulated learning can be classified as discrepancy-reduction models (e.g., Butler & Winne, 1995; Dunlosky & Hertzog, 1997; Dunlosky & Thiede, 1998; Hyland, 1988; Koriat & Goldsmith, 1996; Le Ny, Denhière, & Le Taillanter, 1972; Nelson & Narens, 1990; Powers, 1973; Thiede & Dunlosky, 1999). According to these models, a person begins study by setting a desired state of learning for the to-be-learned material. As the person studies, he or she monitors how well the material has been learned to determine the current state of learning. If the current state of learning meets or exceeds the desired state of learning, the person will terminate study. By contrast, if the current state of learning has not reached the desired state of learning, the person will continue to study the material (by either allocating additional study time to the material or selecting the material for restudy). During restudy, the person monitors learning and compares the current state of learning with the desired state of learning. The person will continue to study until the perceived discrepancy between the current state of learning and the desired state of learning reaches zero.

According to these models, accurate metacognitive monitoring will produce more effective regulation, and this in turn will produce improved learning. However, as noted by Cavanaugh and Perlmutter (1982), there is no strong empirical evidence linking monitoring accuracy or self-regulation to measures of learning. For instance, Begg, Martin, and Needham (1992) examined the relation between monitoring accuracy and test performance. They found that when items were presented once for study, test performance was actually greater for a group of participants that less accurately monitored their learning than for a group that more

accurately monitored their learning. These findings led them to conclude that “memory monitoring does not make a valuable contribution to memory” (p. 212).

Dunlosky and his colleagues (Dunlosky & Connor, 1997; Dunlosky & Hertzog, 1997) have examined the relation between monitoring accuracy and memory by comparing the monitoring accuracy of groups known to differ in memory performance (i.e., older adults vs. younger adults). Groups that differed in performance did not differ in monitoring accuracy. Thus, differences in monitoring accuracy cannot account for the differences in memory performance typically observed between older versus younger adults. In terms of the relation between monitoring accuracy and memory performance, these studies may suggest a weak link between these variables.

Although the aforementioned studies suggest a weak relation between monitoring accuracy and test performance (see also Kelly, Scholnick, Travers, & Johnson, 1976), Maki and Berry (1984) demonstrated that monitoring accuracy was greater for participants who scored above the median on a test than for those who scored below the median, which suggests there is a relation between monitoring accuracy and test performance. Yet, on the basis of a recent review of the literature, Pressley and Schneider (1997) concluded that there is no clear evidence of a relation between prediction (monitoring) accuracy and test performance.

One reason that researchers may have failed to show a relation between monitoring accuracy and test performance is that they have not evaluated the causal relations between these variables. That is, monitoring accuracy may be important because it provides information to guide self-regulation of study, and regulation affects test performance. Thus, the effect of accurate monitoring can only be observed if individuals are free to use this information to regulate their study. When regulation is controlled by the experimenter (as in the investigation by Begg et al., 1992, where study time was held constant across items), the potential influence of monitoring on learning will be minimal. By contrast, one might expect monitoring to affect performance if individuals are allowed to differentially allocate study time on the basis of their monitoring of learning.

The investigations that examined differences in monitoring accuracy between groups known to differ in memory ability are

Keith W. Thiede and Mary C. M. Anderson, Department of Educational Psychology, University of Illinois at Chicago; David Therriault, Department of Psychology, University of Illinois at Chicago.

This research was supported by James S. McDonnell Foundation Grant 98-7 CSEP-EDU.02. Thanks to John Dunlosky for his comments on a draft of this article.

Correspondence concerning this article should be addressed to Keith W. Thiede, Department of Educational Psychology (m/c 147), University of Illinois at Chicago, 1040 West Harrison Street, Chicago, Illinois 60607-7133. E-mail: kthiede@uic.edu

important because they established that differences in test performance might not necessarily be due to differences in accuracy. However, evaluating whether groups that differed in test performance also differed in accuracy is not the same as evaluating whether groups that differed in monitoring accuracy differed in test performance. That is, these investigations did not examine whether monitoring accuracy affects test performance. To do this, it is critical to compare test performance between groups that differ in monitoring accuracy.

A major goal of the present investigation was to evaluate the role of monitoring accuracy in learning. To do this, we used an experimental manipulation that produced different levels of monitoring accuracy. We then examined how differences in accuracy were related to regulation of study (selection of texts for restudy) and in turn how these differences influenced overall learning.

Factors Influencing the Accuracy of Metacognitive Monitoring

Monitoring accuracy can be improved in a variety of ways. For instance, accuracy improves when a person monitors learning after a delay rather than immediately after studying an item (Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991), when items are actively generated during study rather than passively read during study (Mazzoni & Nelson, 1993) and when monitoring occurs after a practice test of the material (King, Zechmeister, & Shaughnessy, 1980; Lovelace, 1984; Shaughnessy & Zechmeister, 1992).

The previous studies involved monitoring learning during an associative learning task. By contrast, in our investigation, we examined monitoring learning during reading of texts. This is not an insignificant methodological difference, given that baseline levels of monitoring accuracy for texts are quite low (Glenberg, Sanocki, Epstein, & Morris, 1987; Maki, 1998) and that most attempts to improve comprehension monitoring have produced less than impressive results (cf. Rawson, Dunlosky, & Thiede, 2000, who achieved relatively high levels of monitoring accuracy by instructing participants to reread texts prior to rating comprehension). For example, Maki and Serra (1992) showed that practice tests had only a modest effect on monitoring accuracy—and only when the practice tests were identical to the eventual tests of comprehension. Unlike this study that showed practice tests have only a small effect on relative accuracy, others have shown that practice tests improve absolute accuracy of monitor (e.g., Ghatala, Levin, Foorman, & Pressley, 1989; Pressley, Snyder, Levin, Murray, & Ghatala, 1987).

On the basis of recent work by Thiede and Anderson (2003), who showed that summarizing texts improved monitoring accuracy, we created differences in monitoring accuracy by instructing participants to generate a list of keywords that captured the essence of a text prior to rating comprehension. That is, participants read six texts. After reading the texts, they wrote a list of five keywords for each text. They then rated their comprehension of each text and took a comprehension test for each text. In a pilot study, generating keywords (vs. not generating keywords) improved monitoring accuracy—however, only when keywords were generated after a delay (not immediately after reading). Thiede and Anderson used activation theories of text comprehension (Britton & Gülgöz, 1991; Fletcher, van den Broek, & Arthur, 1996; van den Broek, Ridsen, Fletcher, & Thurlow, 1996) to explain why the timing of

generation might affect accuracy. According to these theories, spreading activation occurs during reading; thus, more information is active in working memory shortly after reading than after a delay (when activation has decayed). When writing keywords immediately after reading, a person may have access to a highly active mental network. Accordingly, the person may have access to information in short-term memory (STM) to use in writing keywords even for a text that was not well understood. That is, for less understood texts, the person may base keywords on extraneous information activated during reading or on information contained in the text that is active in STM. However, this information (in STM) may not be accessible after the mental network has decayed at the time of the test of comprehension (for a similar discussion of how monitoring information from STM versus long-term memory [LTM] may affect monitoring accuracy of associative memory, see Dunlosky & Nelson, 1992).

The key is that a person may have access to information during keyword generation even for texts that were not well understood, so the process of writing keywords of well-understood texts versus less understood texts may seem quite similar immediately after reading. Therefore, writing keywords immediately after reading may produce a set of homogeneous cues for judging comprehension that may not help discriminate well-understood texts from less understood texts. Moreover, these cues may not be indicative of test performance, given that the test occurs after a delay; therefore, one might expect poor monitoring accuracy when keywords are written immediately after reading. By contrast, activation of the mental network for a text may have decayed for participants when writing keywords after a delay, and a person may have access to only that information retrieved from LTM when writing keywords. Thus, for a less understood text, the person may have little to draw on when writing keywords; whereas, for a well-understood text, the person may retrieve much information during keyword generation. Accordingly, writing keywords after a delay may produce a set of heterogeneous cues for judging comprehension that may highlight differences between well-understood texts and less understood texts. Moreover, these cues are likely highly indicative of test performance because both keyword generation and tests occur after a delay and are based on retrieval of information from LTM; therefore, one might expect high levels of monitoring accuracy when keywords are written after a delay.

In the present experiment, we included three groups. One group wrote keywords after a delay filled by reading other texts (the delayed-keyword group), one wrote keywords immediately after reading a text (the immediate-keyword group), and one did not write keywords (the no-keyword group). The focus of this investigation was on the relation between monitoring accuracy, regulation of study, and test performance. We will evaluate possible explanations for the effect of generating keywords on accuracy elsewhere.

We hypothesized that monitoring accuracy would be greater for the delayed-keyword group than for the other groups. Moreover, we hypothesized that the superior accuracy would lead to more effective regulation of study (i.e., participants would choose to restudy less learned texts over better learned texts to a greater degree) than in the other groups, and this would produce greater test performance for the delayed-keyword group than for the other groups.

Method

Subjects, Design, and Materials

Sixty-six students enrolled in a psychology or educational psychology course at the University of Illinois at Chicago were randomly assigned to three groups (delayed keyword, immediate keyword, or no keyword) by order of appearance. They each received \$20 for their participation.

The texts were seven expository texts taken from encyclopedias on different topics (i.e., communication styles of men vs. women, the effects of alcohol on sleep, experimental design, intelligence and IQ tests, stress, Norse settlements, and World War II naval warfare). The texts ranged in length from 1,118 words to 1,595 words, and ranged in Flesch–Kincaid readability scores from 9.5 to 12.0. A sample text is provided in the Appendix.

Procedure

An overview of the experimental procedure is presented in Figure 1. All participants were instructed that they would read texts, rate their comprehension for each text, and then answer test questions for each text. Participants were also instructed that they might be asked to write a list of keywords that captured the essence of a text. These instructions included an example of keywords (i.e., for a text on the Titanic, one might write *iceberg, shipwreck, tragedy*, etc.), but there was no formal training on how

to generate keywords. All participants, including the no-keyword group, were given pen and paper. Following the instructions, participants were asked to make an ease-of-learning judgment for each text. The ease-of-learning judgment was prompted with the title of the text at the top of the screen and the query “How easily do you think you could learn the information from a passage on the topic listed above? 1 (*very difficult*) to 7 (*very easy*).” After making an ease-of-learning judgment for each text, participants read the sample text, rated their comprehension of the text, and answered the sample questions. Participants were encouraged to ask questions about the procedure during the practice trial.

For the critical trials, the order of text presentation was randomized anew for each participant. Participants in the no-keyword group first read the six texts. After reading, they rated their comprehension for each text. The comprehension rating was prompted with the title of the text at the top of the screen and the query “How well do you think you understood the passage whose title is listed above? 1 (*very poorly*) to 7 (*very well*).” After rating their comprehension of the last text, they answered six questions for each text. Three questions assessed text-based knowledge (details available within a single paragraph of a text), and three inference questions were designed to assess knowledge of a person’s situation model (for a detailed description of various representations of texts and how to assess knowledge of these representations, see Graesser, Millis, & Zwaan, 1997; Kintsch, 1988). Examples of detail and inference questions are provided in the Appendix.

The texts were rated for comprehension and tested in the same order as they were presented for reading. Participants in the delayed-keyword group read each of the six texts. They were then shown the title of a text and instructed to write five keywords that captured the essence of that text. Once they finished writing keywords of the text, they pushed the return key on the computer, which resulted in the presentation of the next title and keyword instructions. After writing keywords for the last text, participants rated their comprehension of each text. After rating their comprehension of the last text, they answered questions for each text. Participants in the immediate-keyword group read a text. They were then shown the title of a text and instructed to write keywords for that text. Once they finished writing keywords for the text, they pushed the return key, which resulted in the presentation of the next text. They read and immediately wrote keywords for each text. After writing keywords for the sixth text, participants rated their comprehension of each text. Following the last comprehension rating, they answered questions for each text.

For each group, after answering the last test question, participants were presented the number of questions they correctly answered over all six tests. That is, they received feedback regarding overall performance but not text specific feedback. Providing explicit text-specific feedback might have reduced the need for an individual to monitor his or her own comprehension. Given the fact that the focus of this investigation was metacognitive monitoring and its effect on regulation, we chose to provide feedback on only overall performance.

Following feedback, participants were then shown an array in which each cell was filled with the title of a text—the cells were numbered from 1 to 6. Participants selected a text for rereading by typing the number of the corresponding cell. After a text had been selected, it was eliminated from the list. They could select zero to six texts for rereading. Texts selected for rereading were randomized anew and presented for rereading. After rereading the final selected text or after selecting no texts for rereading, participants were again tested on each text. This time participants answered 12 questions for each text (6 detail questions and 6 inference questions). Six of these questions were the same as those given on the previous test, and 6 of the questions were new. After responding to the last test question, participants were informed of their performance across all six tests (i.e., total test score).

Procedure		
Instructions		
Rate ease of learning text topics		
No-keyword group	Delayed-keyword group	Immediate-keyword group
Read 1	Read 1	Read 1
Read 2	Read 2	Keyword 1
.	.	Read 2
.	.	Keyword 2
Read 6	Read 6	.
	Keyword 1	.
	Keyword 2	.
	.	.
	.	Read 6
	Keyword 6	Keyword 6
Rate comprehension of the six texts		
Test of the six texts		
Select texts for restudy		
Reread selected texts		
Final test of the six texts		

Figure 1. Overview of the experimental procedure. Read 1 indicates the time that the first text was presented for reading, Read 2 indicates the time that the second text was presented for reading, and so on. Keyword 1 indicates the time that keywords were produced for the first text, Keyword 2 indicates the time that keywords were produced for the second text, and so on.

Results and Discussion

In the present research, all differences reported as reliable are significant at $p < .05$. When interactions were significant and we conducted tests of simple effects, we made a Bonferroni adjustment to maintain a family-wise Type I error rate of .05.

Monitoring Accuracy

As in previous studies (e.g., Glenberg et al., 1987; Maki & Serra, 1992; Weaver, 1990), monitoring accuracy was operationalized as a Goodman–Kruskal gamma correlation between a participant's comprehension rating and initial (prereading) test performance across texts. For each participant, we computed a gamma correlation. The mean of these intraindividual correlations was then computed across participants within each group. Throughout this investigation, we computed a Pearson correlation coefficient whenever we computed a gamma correlation as the measure of association. We chose to report only gamma because this is arguably a more appropriate measure of association, given that it is not affected by an individual's level of test performance or absolute threshold of confidence (for further discussion, see Nelson, 1984). However, please note that throughout this investigation these measures of association led to convergent conclusions.

The mean correlation was reliably greater than zero for all of the groups ($t_s > 2.40$): Comprehension ratings were predictive of subsequent test performance. More important, there was a reliable difference in monitoring accuracy across groups, $F(2, 63) = 4.07$, $MSE = 0.26$. As seen in Figure 2, accuracy was substantially higher for the delayed-keyword group than for the other groups.

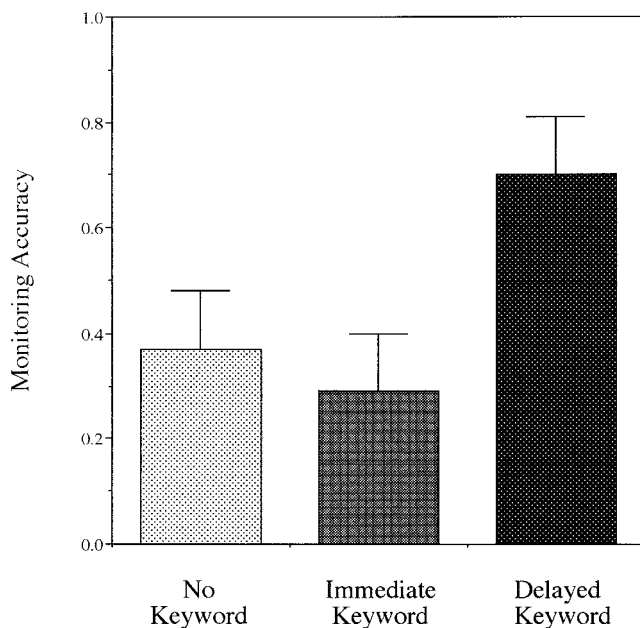


Figure 2. Mean monitoring accuracy presented by group. Monitoring accuracy was operationalized as a Goodman–Kruskal gamma correlation between comprehension ratings and test performance computed across texts for an individual. The data presented are the means of the intraindividual correlations computed across individuals within groups. Error bars represent standard errors of the mean.

Moreover, the accuracy of the delayed-keyword group was as high as any other reported in the literature (only Rawson et al., 2000, and Weaver & Bryant, 1995, showed monitoring accuracy near the level of the delayed-keyword group).

Self-Regulation of Study

According to a discrepancy-reduction model of self-regulated learning, metacognitive monitoring affects learning by influencing regulation of study. The idea is that a person who can accurately discriminate better learned material from less learned material will more effectively regulate his or her study (see Maki, 1995).

Nelson, Dunlosky, Graf, and Narens (1994) and Thiede (1999) showed that test performance was greater for students who allocated more study to material perceived as less learned than to material perceived as better learned. Therefore, we operationalized regulation of study as the correlation between comprehension ratings and whether a text was selected for restudy (where 1 denoted a text had been selected, and 0 denoted a text had not been selected). More effective regulation was indexed by a stronger negative correlation. Given that monitoring accuracy was greater for the delayed-keyword group than for the other groups, we predicted more effective regulation for the delayed-keyword group than for the other groups.

For each participant, we computed a Goodman–Kruskal gamma correlation between a participant's comprehension rating and whether a text had been selected for restudy across texts. The mean of these intraindividual correlations was then computed across participants within each group. As seen in Table 1, participants in the delayed-keyword group were more likely to select less learned texts over better learned texts than were either of the other groups, $F(2, 58) = 3.31$, $MSE = 0.41$. That is, the delayed-keyword group, to a greater degree than the other groups, compensated for initial levels of learning by selecting texts that were less learned for additional study. Given the superior effectiveness of regulated study, we predicted greater performance on the final test of comprehension for the delayed-keyword group than for the other groups. As seen in the right-most column of Table 1, differences in performance cannot be attributed to differences in the number of texts restudied, which did not differ across groups, $F(2, 63) = 1.14$.

Performance on the Second Test of Comprehension

A major goal of this investigation was to evaluate whether monitoring accuracy and self-regulation of study influence test performance. We used a 3×2 analysis of variance to compare performance among the three keyword groups and across two test trials (one before restudy and one after restudy—this was a repeated-measures factor). There was a reliable interaction, $F(2, 63) = 3.27$, $MSE = 0.02$; therefore, we conducted follow-up tests of simple effects. Test performance on the first test was essentially equal for the three groups, $F(2, 63) < 1$, (see Figure 3), which suggests that generating keywords had little impact on initial test performance. However, test performance on the second test, which followed regulation of study, was reliably greater for the delayed-keyword group than for the other groups, $F(2, 63) = 3.90$, $MSE = 0.03$. This difference was due to a substantial increase in performance across trials for the delayed-keyword group,

Table 1
Gamma Correlations Between Comprehension Ratings and Selection of Texts for Restudy (Regulation of Study), and the Number of Texts Selected for Restudy

Group	Regulation of study		Texts selected for restudy	
	<i>M</i>	<i>SEM</i>	<i>M</i>	<i>SEM</i>
Delayed keyword	-.79	.11	2.59	0.20
Immediate keyword	-.35	.17	2.27	0.34
No keyword	-.36	.15	2.86	0.27

$t(21) = 4.07$. By contrast, performance did not change across trials for the immediate-keyword group, $t(21) = 1.59$, and increased only marginally for the no-keyword group, $t(21) = 2.03$.

The superior test performance was consistent across test items that appeared on the first test (old items) and those that were presented for the first time. In particular, the proportion correct on the old items was greater for the delayed-keyword group ($M = .69$, $SEM = .04$) than for the immediate-keyword group ($M = .53$, $SEM = .04$) or the no-keyword group ($M = .56$, $SEM = .05$), $F(2, 63) = 4.11$, $MSE = 0.04$. Likewise, the proportion correct on the new items was greater for the delayed-keyword group ($M = .66$, $SEM = .04$) than for the immediate-keyword group ($M = .51$, $SEM = .05$) or the no-keyword group ($M = .55$, $SEM = .05$), $F(2, 63) = 3.16$, $MSE = 0.04$.

The importance of monitoring accuracy in self-regulated study was indexed by comparison of test performance across groups for texts that were selected for restudy versus those that were not

selected for restudy. By more accurately monitoring comprehension, the delayed-keyword group was better able to identify texts for which they had performed poorly on tests and texts for which they had performed well on tests. They used this information to select texts for restudy. As seen in the two left-most columns of Table 2, for texts selected for restudy, performance on the first test was reliably less for the delayed-keyword group than for the other groups, $F(2, 63) = 3.15$, $MSE = 0.07$. Furthermore, for texts not selected for restudy, performance on the first test was reliably greater for the delayed-keyword group than for the other groups, $F(2, 63) = 8.75$, $MSE = 0.06$.

The superior overall performance on the second test for the delayed-keyword group was produced by dramatically improving performance on texts selected for restudy (see the third column of Table 2) as well as maintaining an existing advantage in test performance for texts not selected for restudy (see the fourth column of Table 2). Test performance on both selected and non-selected texts after participants restudied texts was greater for the delayed-keyword group than the other groups ($F_s > 4.80$).

Conclusion

Accurate metacognitive monitoring is critical to learning. Namely, monitoring provides a basis for making decisions about what to restudy or how long to study material (i.e., regulation of study). Metacognitive monitoring is related to regulation of study (e.g., Mazzoni, Cornoldi, & Marchitelli, 1990; Nelson & Leonesio, 1988), and regulation of study is related to test performance (Thiede, 1999). These studies are important because they provide evidence of the relations suggested by models of self-regulated learning.

By manipulating generation of keywords, we were able to produce groups that varied widely in metacognitive monitoring accuracy. Monitoring accuracy was reliably greater for the delayed-keyword group than for the immediate-keyword group or the no-keyword group. The higher level of accuracy was associated with more effective regulation of study. That is, to a greater degree than the other groups, the delayed-keyword group compensated for initial levels of learning with additional study time. The more effective regulation of study was associated with reliable increases in overall test performance. Thus, this investigation provided empirical evidence to support discrepancy-reduction models of self-regulated learning, with metacognitive monitoring playing an important role in learning.

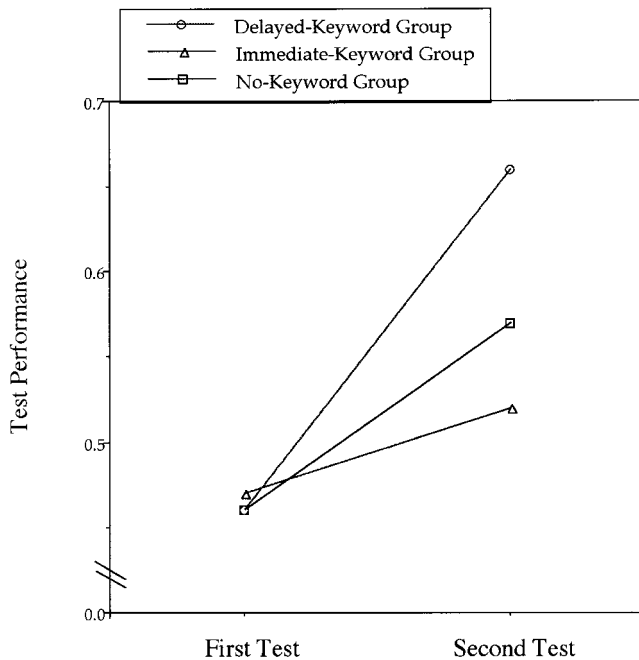


Figure 3. Mean performance by group for the first test (taken prior to rereading selected texts) and the second test (taken after rereading selected texts).

Table 2
Test Performance for Texts That Were Selected Versus Texts That Were Not Selected for Restudy

Group	First test		Second test	
	Selected	Not selected	Selected	Not selected
Delayed keyword	.27	.04	.78	.04
Immediate keyword	.43	.06	.49	.05
No keyword	.44	.06	.55	.06

More accurate monitoring can lead to more effective regulation, which can lead to higher levels of test performance—as was the case with the delayed-keyword group. Lower levels of accuracy, such as those of the immediate-keyword group and the no-keyword group, can lead to less effective regulation of study, which can lead to fairly unimpressive gains in performance resulting from restudying texts.

Given the low levels of accuracy typically reported in metacomprehension literature (for a review, see Maki, 1998), it seems likely that left to their own devices people will not accurately monitor comprehension during reading. If people are not able to distinguish what they understood versus what they did not understand, it seems unlikely that they will differentially allocate study time across material to improve comprehension. Thus, we need to find ways to improve monitoring accuracy as a means of improving reading comprehension.

References

- Begg, I. M., Martin, L. A., & Needham, D. R. (1992). Memory monitoring: How useful is self-knowledge about memory? *European Journal of Cognitive Psychology, 4*, 195–218.
- 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.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*, 245–281.
- Cavanaugh, J. C., & Perlmutter, M. (1982). Metamemory: A critical examination. *Child Development, 53*, 11–28.
- Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition, 25*, 691–700.
- Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multitrial learning. *Journal of Gerontology: Psychological Science, 52*, 178–186.
- Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*, 374–380.
- Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of four factors that affect people's self-paced study. *Acta Psychologica, 98*, 37–56.
- Fletcher, C. R., van den Broek, P., & Arthur, E. J. (1996). A model of narrative comprehension and recall. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 141–164). Mahwah, NJ: Erlbaum.
- Ghatala, E. S., Levin, J. R., Foorman, B. R., & Pressley, M. (1989). Improving children's regulation of their reading PREP time. *Contemporary Educational Psychology, 14*, 49–66.
- Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General, 116*, 119–136.
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48*, 163–189.
- Hyland, M. E. (1988). Motivational control theory: An integrative framework. *Journal of Personality and Social Psychology, 55*, 642–651.
- Kelly, M., Scholnick, E. K., Travers, S. H., & Johnson, J. W. (1976). Relations among memory, memory appraisal, and memory strategies. *Child Development, 47*, 648–659.
- King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology, 93*, 329–343.
- Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review, 95*, 163–182.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490–517.
- Le Ny, J. F., Denhière, G., & Le Taillanter, D. (1972). Regulation of study-time and interstimulus similarity in self-paced learning conditions. *Acta Psychologica, 36*, 280–289.
- Lovelace, E. A. (1984). Metamemory: Monitoring future recall ability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 756–766.
- 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., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 663–679.
- Maki, R. H., & Serra, M. (1992). Role of practice tests in the accuracy of test predictions on text material. *Journal of Educational Psychology, 84*, 200–210.
- Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition, 18*, 196–204.
- Mazzoni, G., & Nelson, T. O. (1993). Metacognitive monitoring after different kinds of monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1263–1274.
- McGovern, T. H. (1993). Norse settlements. In J. E. Cooke (Ed.), *Encyclopedia of the North American colonies* (Vol. 1, pp. 105–107). New York: Scribner.
- Nelson, T. O. (1984). A comparison of current measures of feeling-of-knowing accuracy. *Psychological Bulletin, 95*, 109–133.
- Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: "The delayed JOL effect." *Psychological Science, 2*, 267–270.
- Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5*, 207–213.
- Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 676–686.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–141). New York: Academic Press.
- Powers, W. T. (1973). *Behavior: The control of perception*. Chicago: Aldine.
- Pressley, M., & Schneider, W. (1997). *Introduction to memory development during childhood and adolescence*. Mahwah, NJ: Erlbaum.
- 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.
- Rawson, K., Dunlosky, J., & Thiede, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition, 28*, 1004–1010.
- Shaughnessy, J. J., & Zechmeister, E. B. (1992). Memory monitoring accuracy as influenced by the distribution of retrieval practice. *Bulletin of the Psychonomic Society, 30*, 125–128.
- Thiede, K. W. (1999). The importance of accurate monitoring and effective self-regulation during multitrial learning. *Psychonomic Bulletin & Review, 6*, 662–667.

- Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology, 28*.
- Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1024–1037.
- van den Broek, P., Ridsen, K., Fletcher, C. R., & Thurlow, R. (1996). A “landscape” view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. In B. K. Britton & A. C. Graesser (Eds.), *Models of understanding text* (pp. 165–187). Mahwah, NJ: Erlbaum.
- Weaver, C. A., III. (1990). Constraining factors in calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 214–222.
- 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.

Appendix

Sample Text and Test Questions

The Viking Age Scandinavians had a lasting impact upon the peoples of Western Europe. Their settlements, commercial ventures, and raids affected cultures from the Russian plains to the Irish Sea and from northernmost arctic Norway to the Mediterranean. During the Viking period (ca. 790–1100), Scandinavians also ventured across the North Atlantic, settling the Shetland and Faroe islands, Iceland, and Greenland and making a brief appearance on the shores of America. This North Atlantic arm of the Viking Age expansion connected the eastern and western hemispheres, and, for a few years at the end of the tenth century, a single language and culture reached from Kiev to the gulf of St. Lawrence.

By the beginning of the Viking Age, most of Scandinavia was organized into a maze of local chieftainships. Chieftains were expected to be effective in protecting their clients and aggressive in pressing for every advantage for themselves and their supporters in their struggles with rival chieftains. Traditional law codes (which became increasingly formalized during the Viking period and were written down soon after) and the independence of farmer-clients served somewhat as a restraint on chiefly ambition, but warfare and blood feuds were still commonplace. While Norway, Sweden, and Denmark were known as geographical terms, nothing resembling a nation-state (even by eighth-century standards) existed in pre-Viking Scandinavia.

As wealth from abroad entered Scandinavia, and as Scandinavian merchants, travelers, and mercenaries learned more of the kingdoms of the outside world, the combination of new resources and new ideas seems to have sparked increased competition among local chieftains and petty kings. Agriculture also prospered as a period of warm climate (now known as the Little Climatic Optimum) lengthened growing seasons in northwestern Europe. Population seems to have enlarged, which led to the settlement of the uplands and the extension of Norse farms into arctic Norway. The expansion of territorial boundaries during the Viking Age provided an outlet for this growing rural population and yielded new territory for the losers in the intensifying struggles among chieftains for dominance.

Neither a growing population nor competing chieftains would have produced the Viking expansion had the means for overseas travel, trade, and conquest been lacking. Through the efforts of maritime archaeologists, we know a good deal about Viking period ships and their construction. By the late eighth century, Scandinavian clinker-built ships had reached a high level of perfection, combining lightness and shallow draft with great strength and sea-keeping ability. Viking ships could land on any beach, penetrate far up rivers, and survive North Atlantic storms on the open sea.

While strong and elegant, the clinker-built Viking ships had two significant limitations. They required a long run of high-quality timber (preferably oak) for the keel and naturally curved timbers for the stem and stern pieces. Since this quality timber was absent in the North Atlantic islands, settlers in Iceland and Greenland found it hard to replace oceangoing ships lost at sea. The Viking design also sharply limited cargo capacity—even the knarrs (trading vessels) could carry only a fraction of the cargo of the

later carvel-built Hanseatic cogs that came to dominate European commerce in the later Middle Ages. Viking ships could reach distant points, but they could not carry enough passengers and supplies to ensure a viable transatlantic foothold. Population movement across the North Atlantic thus required a chain of settlements, each providing population and resources for successive ventures westward.

Scandinavian North Atlantic settlement was a gradual process taking two hundred years to complete. Norse colonists settled the Shetlands and Orkneys around the year 800 and (according to tradition) Iceland around AD 874. Greenland was settled from Iceland by Eirik the Red around 985. Vinland was explored from Greenland and a settlement was attempted by the sons and daughter of Eirik around the year 1000.

Island chieftains who (like Eirik) had failed in local power struggles provided the ships and capital to sponsor further voyages of exploration and settlement. Unsuccessful farmers and dissatisfied younger siblings from successively filled island ecosystems provided the bulk of the personnel. The first settlers in a new land had the ritually important right to name the landscape and economically vital right to claim the best pasture and hunting grounds. As prime grazing is often patchy and limited in the North Atlantic islands, this initial division of resources set the stage for increasing economic and social hierarchy in later generations.

During the eleventh and twelfth centuries, the Scandinavian North Atlantic enjoyed modest prosperity. Island populations seem to have stabilized at low levels; Iceland’s population was probably between thirty thousand and sixty thousand, and Greenland’s was six thousand at most. While state formation was taking place in the Scandinavian homelands, the more distant North Atlantic islands seem to have maintained a somewhat archaic chiefly oligarchy. Christianity had spread as far as Greenland by the year 1000, and most Scandinavians were at least nominally Christian by 1100. Chiefly competition was now conducted through the endowment of churches and monastic houses as well as by the traditional sheep stealing and house burning. In Iceland and probably Greenland, sagas and family histories were being composed, and poets and skalds from the North Atlantic were still in demand in continental courts.

Along with prosperity came the beginnings of decline. Iceland’s chiefly dominance struggles had thrown up six great families whose escalating warfare increasingly exhausted local resources. Overgrazing in many areas triggered massive and irreversible soil erosion, turning whole districts into rocky wasteland. After 1250, volcanic eruptions coupled with the end of the favorable weather of the Little Climatic Optimum added to man-made disaster, and increasing numbers of North Atlantic farmers slipped from freeholder to tenant status.

After 1264, Iceland and Greenland became part of the Norwegian kingdom just as that kingdom was about to enter a long period of decline. Their local oceangoing ships long lost, the settlers of the western Atlantic depended upon continental merchants to carry their trade. Icelanders bitterly complained that the promised six ships per year seldom arrived, and

it seems to have taken a papal letter five years to reach Greenland. The eastern Atlantic settlements in the Shetlands and northern Scotland were luckier, as they were becoming increasingly integrated into the stock fish trade through the Hanseatic League.

The late thirteenth and the fourteenth centuries saw accelerated decline in the western North Atlantic. The onset of the Little Ice Age (ca. 1250–1860 in the North Atlantic) crippled farming, and economic hardship in Norway affected transatlantic trade. Literature declined, and the populations of Iceland and Greenland became locked in a struggle for bare survival. By the later Middle Ages, the Norse North Atlantic was no longer the cutting edge of an expanding European population but a demoralized and isolated backwater.

Detail question: Which was settled first by Norse colonists?

- A. Greenland
- B. Iceland
- C. The Shetlands*
- D. Vinland

Inference question: What modern group is most like the chieftainships of the Viking period?

- A. Urban gangs*
- B. Midwest farmers
- C. State legislators
- D. College students

Note. Asterisks indicate correct answer. Text (excluding questions) from "Norse Settlements," by T. H. McGovern, in *Encyclopedia of the North American Colonies* (Vol. 1, pp. 105–107), edited by J. E. Cooke, 1993, New York: Charles Scribner's Sons. Copyright 1993 by Charles Scribner's Sons. Reprinted by permission of The Gale Group.

Received July 24, 2001
 Revision received March 8, 2002
 Accepted March 14, 2002 ■



**AMERICAN PSYCHOLOGICAL ASSOCIATION
 SUBSCRIPTION CLAIMS INFORMATION**

Today's Date: _____

We provide this form to assist members, institutions, and nonmember individuals with any subscription problems. With the appropriate information we can begin a resolution. If you use the services of an agent, please do **NOT** duplicate claims through them and directly to us. **PLEASE PRINT CLEARLY AND IN INK IF POSSIBLE.**

PRINT FULL NAME OR KEY NAME OF INSTITUTION	MEMBER OR CUSTOMER NUMBER (MAY BE FOUND ON ANY PAST ISSUE LABEL)
ADDRESS	DATE YOUR ORDER WAS MAILED (OR PHONED)
CITY STATE/COUNTRY ZIP	<input type="checkbox"/> PREPAID <input type="checkbox"/> CHECK <input type="checkbox"/> CHARGE CHECK/CARD CLEARED DATE: _____
YOUR NAME AND PHONE NUMBER	(If possible, send a copy, front and back, of your cancelled check to help us in our research of your claim.) ISSUES: <input type="checkbox"/> MISSING <input type="checkbox"/> DAMAGED

TITLE	VOLUME OR YEAR	NUMBER OR MONTH

Thank you. Once a claim is received and resolved, delivery of replacement issues routinely takes 4–6 weeks.

(TO BE FILLED OUT BY APA STAFF)	
DATE RECEIVED: _____	DATE OF ACTION: _____
ACTION TAKEN: _____	INV. NO. & DATE: _____
STAFF NAME: _____	LABEL NO. & DATE: _____

Send this form to APA Subscription Claims, 750 First Street, NE, Washington, DC 20002-4242

PLEASE DO NOT REMOVE. A PHOTOCOPY MAY BE USED.