The importance of monitoring and self-regulation during multitrial learning

KEITH W. THIEDE
University of Illinois, Chicago, Illinois

Theory suggests that accuracy of metacognitive monitoring and self-regulation of study will affect test performance, but there is little empirical evidence linking these variables. I examined the relation among these variables in a multitrial learning task. Regression analyses showed that monitoring accuracy and self-regulation were reliably related to test performance—greater monitoring accuracy and more effective self-regulation were associated with greater test performance. These analyses were contrasted with analyses typically conducted in previous research, to show the importance of using a multitrial learning task and of attending to the theoretically based causal relation among variables when evaluating how monitoring accuracy and self-regulation are related to test performance. The results of this investigation may help to explain why previous research has failed to link these variables.

Recently, a great deal of research has focused on metacognitive monitoring and self-regulation of study (for books highlighting research in these areas, see Hacker, Dunlosky, & Graesser, 1998; Metcalfe & Shimamura, 1994; Reder, 1996; Schunk & Zimmerman, 1994). The interest in these topics may be due to the conceptual appeal of the proposition that accurate monitoring and self-regulation of study play an important role in learning.

Consider a metacognitive perspective on the learning process. As a person studies, he or she monitors how well material has been learned. The output from monitoring is then used as a basis for regulating study (e.g., deciding whether to continue to study the material). Eventually, the person will try to retrieve the material from memory. On the basis of this model, accurate monitoring of learning and self-regulation of study will produce higher performance on a test of memory. For instance, if a person accurately monitors what he or she has learned and not learned, the person can allocate more study to less well learned material, which should increase learning of that material and improve overall test performance. However, there is no strong empirical evidence for this link between monitoring accuracy or self-regulation and measures of learning (Cavanaugh & Perlmutter, 1982).

Accurate Monitoring and Test Performance

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. When items were presented twice for study, test performance did not differ between the groups. On the basis of these findings, Begg et al. concluded that "memory monitoring does not make a valuable contribution to memory" (p. 212).

Others have also found no relation between monitoring accuracy and test performance (Kelly, Scholnick, Travers, & Johnson, 1976). However, Maki and Berry (1984) showed that monitoring accuracy was greater for participants who scored above the median on a test than for participants who scored below the median on the test, suggesting 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.

Self-Regulation and Test Performance

What is effective self-regulation of study? One might argue that effective self-regulation involves allocating more study time to more difficult items than to less difficult items (i.e., compensating for the difficulty of items with additional study time). Indeed, this is a commonly observed pattern of allocating study time (e.g., Dunlosky & Connor, 1997; Mazzoni & Cornoldi, 1993; Mazzoni, Cornoldi, & Marchitelli, 1990; Nelson & Leonesio, 1988).

Regarding the relation between self-regulation and test performance, one research strategy has been to compare self-regulation among groups of individuals known to differ in memory ability. Dunlosky and Hertzog (1997) found no difference in self-regulation among older and younger adults, and Masur, McIntyre, and Flavell (1973) found no differences in self-regulation among third grade students and college students, even though substantial age-related differences were found in test performance. Cull and Zechmeister (1994) found no difference in self-regulation among more successful and less successful

learners (level of success was operationalized as performance on a practice test), even though these groups differed in performance. Cull and Zechmeister concluded that performance differences among more successful and less successful learners "have more to do with learning effectiveness than with metamemory" (p. 256). That is, because groups that differ in performance regulate study in similar ways, self-regulation cannot account for the differences in performance.

Causal Relations Among Monitoring Accuracy, Self-Regulation, and Learning

One reason that researchers may have failed to show that test performance is affected by monitoring accuracy and self-regulation is that the causal relations among these variables have not been evaluated. Theory suggests that monitoring is important because it provides information that guides self-regulation of study and these decisions about what to study and how long to study affect 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 self-regulation of study among groups known to differ in memory ability are important because they showed that differences in test performance may not necessarily be due to differences in self-regulation. However, these investigations did not empirically evaluate the causal relation between self-regulation and test performance. In particular, these studies evaluated whether groups that differ in test performance also differ in self-regulation, but not whether groups that differ in self-regulation also differ in test performance. To examine the causal relation, one must compare performance among groups known to differ in self-regulation.

Nelson, Dunlosky, Graf, and Narens (1994) conducted an experiment to evaluate the effect of regulation of study on performance in a multitrial learning task. In this experiment, a computer selected items for restudy (half the items of the list). In one condition, the computer selected for restudy those items judged as least learned, which would be an effective selection strategy, given that additional study would compensate for initial judged item difficulty. In another condition, the computer selected for restudy those items judged as best learned, which would be an extremely ineffective selection strategy. In yet another condition, participants self-selected items for restudy. In each of these conditions, the items selected on the first trial were presented on each of the subsequent study trials. Nelson et al. found that test performance was greater for the group of participants who restudied the least

learned items and for those who restudied self-selected items than for the group of participants who restudied the best learned items.

Atkinson (1972, 1976) also showed that decisions about how to regulate study affect test performance. In particular, he showed that test performance was greater for participants who self-selected items for study and for participants whose item selection was based on previous recall performance (items not previously recalled were presented for study) than for participants who studied a set of randomly selected items.

The investigations by Nelson et al. (1994) and Atkinson (1972, 1976) showed that effective regulation versus ineffective regulation produced differences in test performance. However, in these investigations, the ineffective regulation groups may have been too ineffective to serve as realistic comparison groups: People typically use more effective strategies than randomly selecting items for study (see, e.g., Mazzoni et al., 1990; Nelson & Leonesio, 1988) and far more effective ones than selecting the best learned items of the list—the extremely ineffective strategy for the regulation group in Nelson et al. Thus, it has yet to be shown how regulation of study within a typical range of effectiveness is related to test performance. Moreover, in Nelson et al., decisions about which items to restudy were held constant across trials (i.e., the items selected for restudy after the first study trial were presented for restudy on all the remaining study trials); therefore, this study did not investigate how regulation may change across trials and how this may affect learning.

The focus of this investigation was on the question of whether participants monitor the degree of learning for one item relative to another item and whether they use this information to differentially allocate study time across items of a list. To measure this kind of monitoring, I had participants make a judgment of learning (JOL) for each item, which is a prospective judgment of the likelihood of recalling an item on an eventual test. The accuracy of JOLs can be inferred from the correlation between JOLs and actual test performance—a higher positive correlation indicates higher monitoring accuracy. Thus, in contrast to examining the accuracy of list-level predictions (called *absolute accuracy*; see Levin, Yussen, De Rose, & Pressley, 1977; Pressley & Ghatala, 1989), I examined the accuracy of item-by-item predictions (called *relative accuracy*). This relative accuracy is important because it provides a measure of whether people judge which items are known versus not known, which is essential information if people are to regulate their study effectively—that is, allocate more study to less well learned items than to better learned items (Maki & Serra, 1992). The effectiveness of self-regulation can be inferred from the correlation between JOLs and item selection. A strong negative correlation between JOLs and item selection indicates a greater degree of compensation for initial item difficulty (i.e., more effective self-regulation). I examined whether monitoring accuracy and self-regulation were related to test performance on a given trial.

Table 1
Mean Recall, Judgment of Learning, Number of Items Selected for
Restudy, Judgment of Learning (JOL) Accuracy, and Self-Regulation

Phase	Recall		JOL		Number of Items Selected for Restudy		JOL Accuracy		Self-Regulation	
	M	SE	M	SE	M	SE	M	SE	M	SE
Accuracy										
Trial 1	.31	.01	16.6	2.0			.76	.03		
Regulation										
Trial 1	.51	.02	23.0	2.7	25.3	0.9			−.74	.05
Trial 2	.64	.03	40.9	3.5	20.5	1.0			−.70	.05
Trial 3	.75	.03	53.0	3.8	18.0	1.1			−.71	.06
Trial 4	.78	.03	52.8	4.3	16.3	1.2			−.53	.09
Trial 5	.80	.03	49.6	4.6	14.3	1.4			−.41	.13

Note—JOL accuracy was operationalized as a Goodman–Kruskal gamma correlation between JOLs and test performance. Self-regulation was operationalized as a Goodman–Kruskal gamma correlation between JOLs and selection of items for restudy.

METHOD

Participants, Materials, Apparatus, and Design

The participants were 120 students from the University of Illinois at Chicago, who participated individually for extra class credit. Items consisted of the 36 Swahili–English translation equivalents (e.g., *ardhi*–soil) used by Nelson et al. (1994), which were selected from a list normed by Nelson and Dunlosky (1994). A Macintosh computer displayed all instructions and controlled all aspects of the procedure. The computers also recorded participants' responses. The method of presenting items for selection for restudy was manipulated between subjects (array vs. item by item).

Procedure

The participants were instructed that their goal was to learn all the items of the list and that the experiment would end when they correctly recalled all the items on a test trial or after they completed six test trials. The instructions described and provided examples of all aspects of the procedure. The participants were encouraged to ask questions during the instructions.

Following the instructions, the items were randomly ordered and presented for study. For study, items were individually presented for 4 sec/item. After the entire list had been presented, the participants made a JOL for each item (i.e., a delayed JOL; see Nelson & Dunlosky, 1991). JOLs were prompted with only the stimulus of an item (e.g., if "*ardhi*–soil" had been presented for study, the JOL cue would be "*ardhi*–?") and the query, "How confident are you that in about ten minutes from now you will be able to recall the second word of the pair when prompted with the first?" (0 = *definitely won't recall*, 20 = *20% sure*, 40..., 60..., 80..., and 100 = *definitely will recall*).

After a JOL had been made for each item, the items were presented for paired-associate recall. For each recall trial, the stimulus of an item was presented, and the participants were instructed to type the correct response. Omissions were not allowed. After the final item had been presented for paired-associate recall, the number of items correctly recalled across the 36 items was presented.

The above study–JOL–test trial provided data for measuring the accuracy of JOLs (i.e., the correlation between JOLs and test performance). This accuracy phase was followed by a regulation phase, in which the participants regulated their study by selecting items for restudy. In this phase, the participants again made a JOL for each item. They then selected items for restudy. As in Thiede and Dunlosky (1999), the participants selected items in one of two ways. In the array condition, the participants were shown a 3 × 12 array in

which each cell was filled by a stimulus of a different item of the list; the cells were numbered from 1 to 36. The participants selected an item for restudy by typing the number of the corresponding cell. After an item had been selected, it was eliminated from the array. The participants could select 0 to 36 items. In the item-by-item condition, during selection, the stimulus of an item was presented with the query, "Do you want to restudy this item? Yes = 1, No = 0." The participants went through the list item by item; they could select 0 to 36 items for restudy.

Items selected for restudy were randomized anew and presented for restudy (4 sec/item). Presentation of the last item selected for restudy was followed by paired-associate recall. The number of items correctly recalled was presented after the final item was presented for test. If all the items were correctly recalled, the experiment ended; otherwise another JOL–selection–restudy–test trial occurred. This procedure continued until either all the items were correctly recalled on a trial or six test trials had been completed.

RESULTS AND DISCUSSION

Measures

Test performance. Test performance was measured on each of six trials or until a participant mastered the list. The first test trial provided a measure of initial performance and occurred before the participants selected items for study (the accuracy phase). The next five tests reflected performance potentially influenced by additional study (the regulation phase). For the participants who mastered the list before the fifth test of the regulation phase (sixth test, overall), a perfect test score was entered for each trial following the test that showed that the list had been mastered. As can be seen in Table 1, recall performance increased across trials.

Judgments of learning. JOLs were measured on each of six trials or until a participant mastered the list. For each trial, the median JOL was computed for each participant. The mean of the medians was then computed across participants. As can be seen in Table 1, mean JOL increased across the early trials and then leveled off over the last three trials. This plateau may have been due to missing data. For those participants who mastered the list prior to

Table 2
Summary of Regression Analysis for Variables
Predicting Test Performance by Regulation Trial

Variable	Trial 1 ($N = 96$)			Trial 2 ($N = 95$)			Trial 3 ($N = 84$)			Trial 4 ($N = 56$)			Trial 5 ($N = 32$)		
	B	$SE\ B$	β	B	$SE\ B$	β	B	$SE\ B$	β	B	$SE\ B$	β	B	$SE\ B$	β
Accuracy	.19	.11	.17	.35	.11	.28*	.19	.09	.18*	.28	.10	.29*	.25	.12	.25*
Regulation	−.11	.04	−.28*	2.24	.05	−.47*	−.29	.04	−.62*	−.20	.05	−.48*	−.18	.04	−.49*
Number selected	.00	.01	.04	.00	.01	−.15	.00	.01	−.16	.01	.01	−.18	.00	.01	−.10

Note—$R^2 = .13$ for Trial 1; $R^2 = .33$ for Trial 2; $R^2 = .46$ for Trial 3; $R^2 = .47$ for Trial 4; $R^2 = .39$ for Trial 5. *$p < .05$.

the last trial (29 and 38 participants Trials 4 and 5 of the regulation phase, respectively), trials following mastery had missing data for JOLs. Thus, the mean JOLs on later trials may have been lower because those means were based on JOLs of those participants who had not mastered the list of items.

Number of items selected for restudy. Item selection was measured on the five trials of the regulation phase or until a participant mastered the list. For those participants who mastered the list prior to the last trial, trials following mastery had missing data for item selection. The number of items selected for restudy decreased across trials (see Table 1).

Accuracy of judgments of learning. Monitoring accuracy is a measure of how well JOLs predict subsequent recall. Monitoring accuracy was operationalized as a Goodman–Kruskal gamma correlation (Goodman & Kruskal, 1954) between JOLs and recall on the test trial occurring prior to item selection and restudy (i.e., the test occurring during the accuracy phase). For a rationale for using gamma as a measure of association between measures such as those used in this investigation, see Nelson (1984). Each participant had one measure of JOL accuracy. As in previous investigations (Nelson & Dunlosky, 1991), these delayed JOLs correlated quite highly with test performance. The mean gamma, computed across participants, was +.76 ($SEM = .03$); the median was +.84. The positive correlation indicates that items given higher JOLs were more likely to be recalled than were items given lower JOLs—only 4 participants had a gamma below zero (indicating extremely inaccurate predictive accuracy).

Judgments of learning and recall. A measure of JOL accuracy was available only on the first trial. On subsequent trials (the regulation phase), the participants may have compensated for lower levels of learning by allocating study time to those items given lower JOLs. Differential allocation of study time may make JOLs less predictive of subsequent recall, thereby weakening the relation between JOLs and subsequent recall (see the monitoring-neutralization hypothesis; Nelson & Leonesio, 1988). Although this is not central to this paper, I evaluated this hypothesis by computing two gamma correlations for each trial of the regulation phase. First, I computed a Goodman–Kruskal gamma correlation between JOLs and subsequent recall. Next, I computed a Goodman–Kruskal gamma correlation between JOLs and prior recall. The

monitoring-neutralization hypothesis suggests that JOLs will be less predictive of recall if individuals regulate study to compensate for differences in judged learning. That is, JOLs are expected to be less predictive of subsequent recall than of prior recall.

The mean gamma correlation between JOLs and subsequent recall on Trial 1 of the regulation phase was +.74 ($SEM = .03$). On subsequent trials, the correlations were .62, .58, .62, and .61, respectively. The mean gamma correlation between JOLs on Trial 1 and recall on the previous trial (recall in the accuracy phase) was +.88 ($SEM = .02$). On subsequent trials, the relation was also strong and positive (.87, .85, .81, and .67, respectively). These latter correlations were marginally greater than the former correlations [$F(1,14) = 3.73, p = .07$], which provides some support for the monitoring-neutralization hypothesis.

The high correlations between JOLs and prior recall may be seen as support for the hypothesis that JOLs are based on cues resulting from an attempt to recall an item (as was suggested by King, Zechmeister, & Shaughnessy, 1980), but the lack of a perfect correlation and a binomial distribution of JOLs suggests that JOLs are not based entirely on previous recall.

Self-regulation of study. Self-regulation was operationalized as a Goodman–Kruskal gamma correlation between JOLs and item selection on the five trials of the regulation phase. Item selection was coded 1 for items selected for restudy and 0 for items not selected for restudy; thus, a strong negative correlation indicated a greater degree of compensation for levels of perceived learning. Each participant had, at most, five measures of self-regulation. For Trials 4 and 5 of the regulation phase, more than half the cases were excluded owing to missing data for self-regulation (64 and 72 participants for Trials 4 and 5 of the regulation phase, respectively). The participants had missing data partly because of indeterminant gamma correlations for self-regulation resulting from invariance in item selection or JOLs. They also had missing data as a result of missing item selection data, owing to mastering the list prior to the last trial. Excluding cases on the basis of mastery of the list may be problematic, because the most effective learners were systematically dropped; however, as can be seen in Table 1, the measures of self-regulation on the latter trials are consistent with the measures on earlier trials. The strong negative correlations indicate that, across trials, the par-

ticipants were more likely to choose to restudy items given lower JOLs over those given higher JOLs.

Are Monitoring Accuracy and Self-Regulation of Study Related to Test Performance?

A major goal of this investigation was to evaluate whether monitoring accuracy and self-regulation of study influence test performance. For each of the five regulation trials, a standard regression analysis was conducted—test performance was regressed on monitoring accuracy and self-regulation. In addition, because test performance may simply be a function of the number of items selected for restudy and not so much of which items were selected for restudy, as is indicated by the measure of self-regulation, the number of items selected for restudy for a particular trial was also included in the regression analysis (the predictor variables were simultaneously entered).

Although monitoring accuracy was not reliably related to test performance on trial 1 ($p = .11$), it was reliably related to test performance on later trials (see Table 2). Self-regulation was reliably related to test performance on all the trials. By contrast, the number of items selected for restudy was not related to test performance, which indicates that test performance was more a function of which items a person chose to restudy than of how many items a person chose to restudy. These results provide empirical evidence linking monitoring accuracy and self-regulation to test performance. Greater monitoring accuracy (a stronger positive correlation between JOLs and recall) and more effective self-regulation (a stronger negative correlation between JOLs and item selection) were associated with greater test performance.

It is particularly important to note that monitoring accuracy was not related to performance on the first trial, given that previous investigations evaluating the relation between accuracy and performance have used single-trial learning tasks. The findings above suggest that the effects of monitoring accuracy on test performance may only be observed after multiple study trials.

In contrast to the above findings, consider the results of analyses typically conducted to examine the effects of monitoring accuracy and self-regulation on test performance (i.e., comparing accuracy and self-regulation among groups known to differ in test performance). I performed a median split on performance on the first test, classifying participants who scored above the median as more successful learners and those scoring below the median as less successful learners (as in Cull & Zechmeister, 1994). Monitoring accuracy did not differ for more successful learners ($M = +.78$, $SEM = .04$) versus less successful learners [$M = +.74$, $SEM = .05$; $t(110) < 1$]. Moreover, self-regulation did not differ for more successful learners ($M = -.75$, $SEM = .07$) versus less successful learners ($M = -.72$, $SEM = .07$) on the first trial [$t(96) < 1$], nor did it differ for groups across the trials [$F(4,68) = 0.41$, $MS_e = 0.18$]. On the basis of these findings, one may be tempted to conclude that monitoring accuracy and self-regulation do not affect test performance,

which is in contrast with the conclusion drawn from results of the prior analyses. Clearly, evaluating whether groups that differ in test performance differ in accuracy or self-regulation is different from evaluating whether groups that differ in accuracy or self-regulation differ in test performance. This may help to explain why previous research has failed to find a relation between these variables.

Conclusions

The findings of this investigation provide the first empirical evidence to rebut claims that monitoring accuracy and self-regulation do not contribute to learning. Monitoring accuracy and self-regulation of study were related to test performance (see Table 2). As is predicted by a metacognitive model of learning, more accurate monitoring and more effective self-regulation of study were associated with better test performance.

The present investigation also suggests the importance of attending to theoretically derived causal relations when evaluating the effect of monitoring and self-regulation on test performance. Looking for differences in performance among groups that differ in monitoring accuracy or self-regulation is not the same as looking for differences in accuracy and self-regulation among groups that differ in performance. This was evident from the conflicting conclusions drawn from two analyses of the same data. In particular, the analysis of the data that did not attend to the causal relations among the variables (evaluating whether groups that differ in performance also differ in monitoring accuracy and self-regulation) led to the conclusion that monitoring accuracy and self-regulation were not related to test performance. This is in sharp contrast with the conclusions drawn from the more causal-based analysis, which showed reliable relations among these variables.

Failing to attend to causal relations among variables may explain why previous research has failed to show that monitoring accuracy and self-regulation are related to test performance. Another reason may be that previous investigations have used single-trial learning tasks, rather than multitrial learning tasks. In this investigation, the relation between monitoring accuracy and test performance was not evident until after the first restudy trial. Thus, just as Nelson et al. (1994) showed that the effects of effective regulation on performance increased across trials, the present results suggest that the effects of monitoring accuracy may also increase across trials.

Implications for Education

Research on metacognitive training has tended to focus on interventions designed to increase student's ability to monitor the effectiveness of different study strategies (e.g., Ghatala, Levin, Pressley, & Lodico, 1985; Lodico, Ghatala, Levin, Pressley, & Bell, 1983). The results of the present investigation show that more accurate monitoring (discriminating between what is known vs. not known) is associated with increased learning. Thus, a new focus for metacognitive training may be to teach students

to discriminate between what they know versus what they don't know. Increasing monitoring accuracy may improve learning via its impact on the self-regulation of study.

REFERENCES

ATKINSON, R. C. (1972). Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology*, **96**, 124-129.

ATKINSON, R. C. (1976). Adaptive instructional systems: Some attempts to optimize the learning process. In D. Klahr (Ed.), *Cognition and instruction* (pp. 81-108). Hillsdale, NJ: Erlbaum.

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.

CAVANAUGH, J. C., & PERLMUTTER, M. (1982). Metamemory: A critical examination. *Child Development*, 53, 11-28.

CULL, W. L., & ZECHMEISTER, E. B. (1994). The learning ability paradox in adult metamemory research: Where are the metamemory differences between good and poor learners? *Memory & Cognition*, **22**, 249-257.

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.

GHATALA, E. S., LEVIN, J. R., PRESSLEY, M., & LODICO, M. G. (1985). Training cognitive-strategy monitoring in children. *American Educational Research Journal*, **22**, 199-215.

GOODMAN, L. A., & KRUSKAL, W. H. (1954). Measures of association for cross-classification. *Journal of the American Statistical Association*, **49**, 732-764.

HACKER, D. J., DUNLOSKY, J., & GRAESSER, A. C. (Eds.) (1998). *Metacognition in educational theory and practice*. Mahwah, NJ: Erlbaum.

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.

LEVIN, J. R., YUSSEN, S. R., DE ROSE, T. M., & PRESSLEY, M. (1977). Developmental changes in assessing recall and recognition memory capacity. *Developmental Psychology*, **13**, 608-615.

LODICO, M. G., GHATALA, E. S., LEVIN, J. R., PRESSLEY, M., & BELL, J. A. (1983). The effect of strategy-monitor training on children's selection of effective memory strategies. *Journal of Experimental Child Psychology*, **35**, 263-277.

MAKI, R. H., & BERRY, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, & 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.

MASUR, E. F., MCINTYRE, C. W., & FLAVELL, J. H. (1973). Developmental changes in apportionment of study time among items in a multitrial free recall task. *Journal of Experimental Child Psychology*, **15**, 237-246.

MAZZONI, G., & CORNOLDI, C. (1993). Strategies in study time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: General*, **122**, 47-60.

MAZZONI, G., CORNOLDI, C., & MARCHITELLI, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition*, **18**, 196-204.

METCALFE, J., & SHIMAMURA, A. P. (1994). *Metacognition: Knowing about knowing*. Cambridge, MA: MIT Press.

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. (1994). Norms of paired-associate recall during multitrial learning of Swahili–English translation equivalents. *Memory*, **2**, 325-335.

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, & Cognition*, **14**, 676-686.

PRESSLEY, M., & GHATALA, E. S. (1989). Metacognitive benefits of taking a test for children and young adolescents. *Journal of Experimental Child Psychology*, **47**, 430-450.

PRESSLEY, M., & SCHNEIDER, W. (1997). *Introduction to memory development during childhood and adolescence*. Mahwah, NJ: Erlbaum.

REDER, L. M. (1996). *Implicit memory and metacognition*. Mahwah, NJ: Erlbaum.

SCHUNK, D. H., & ZIMMERMAN, B. J. (1994). *Self-regulation of learning and performance*. Hillsdale, NJ: Erlbaum.

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, & Cognition*, **25**, 1024-1037.