Memory & Cognition
2007, 35 (4), 668-678

The negative cascade of incongruent generative study–test processing in memory and metacomprehension

Ayanna Kim Thomas
Tufts University, Medford, Massachusetts

and

Mark A. McDaniel
Washington University, St. Louis, Missouri

Previous research suggests that when participants engage in generative study activities, the processing of text is enhanced and improvements in memory and metacomprehension result. However, few studies have investigated the influence of processes required by the testing situation or the interaction between encoding and retrieval processes on metacomprehension accuracy. The present experiments examine whether the congruency of processes generated during study and required at retrieval affect memory, metacomprehension, and control processes. Study orientation and test type were congruent (i.e., letter-reinsertion: detailed test), incongruent (i.e., letter-reinsertion: conceptual test), or neutral (i.e., read: conceptual test). After generative study, but before testing, participants made metacomprehension predictions for previously studied texts. Controlled strategy selection was measured in Experiment 2. When processes at study and test were congruent, cued recall performance and metacomprehension predictions were more accurate than when study and test were incongruent. For incongruent conditions, metacomprehension predictions were no better than chance; thus, controlled strategy selection was based on inaccurate metacomprehension, thereby further penalizing memory performance relative to congruent conditions. These findings extend a transfer-appropriate processing framework to metacomprehension.

Integral to the process of reading is one's ability to assess whether the read information has been comprehended and can later be recalled. This introspective process requires that we, as learners, consciously evaluate various components of learning to decide whether we have successfully mastered the material (Flavell, 1979). Consider, for example, the student preparing for an exam. Ideally, the student sets a desired state of mastery of the material and regularly monitors to see whether that desired state has been reached. If the desired state has not yet been reached, the student implements additional controlled processes (e.g., allocating additional study time) to master the material. Monitoring the comprehension of to-be-learned material, described as metacomprehension, can include judgments about levels of comprehension and learning of the text, predictions about future memory for the material, and the ability to correct oneself while reading (Rudell & Speaker, 1985). As predictions about future memory become more accurate, the reading process can become more efficient.

The present study examines conditions that may improve predictions about future performance. Preliminary studies have identified generative encoding activities as important factors in metacomprehension accuracy. We extend these findings by investigating the effect of two generative activities on metacomprehension within a transfer appropriate processing (TAP) framework (McDaniel, Friedman, & Bourne, 1978; Morris, Bransford, & Franks, 1977). According to TAP, the effectiveness of generative study activities will depend on the extent to which the processes required by the study activity correspond to those required at retrieval. As one example, Nairne and Widner (1987) demonstrated that test appropriateness mediated the generation effect in nonwords. In two experiments, we confirm the predictions made by the TAP framework, using stimuli and memory tests reflective of educational settings. More importantly, for the first time, we examine the influence of TAP on metacomprehension by investigating whether metacomprehension predictions and study-time allocation are affected by the relationship between processing at study and processing at test. Before reporting the experiments, we first develop the pertinent empirical and theoretical background.

Early investigations (i.e., Glenberg & Epstein, 1985; Glenberg, Sanocki, Epstein, & Morris, 1987; Glenberg, Wilkinson, & Epstein, 1982; Maki & Berry, 1984) into metacomprehension reported only small or no relationships between predictions of future performance and later

A. K. Thomas, ayanna.thomas@tufts.edu

recall of text material. However, more recently, researchers have demonstrated that participants can accurately predict future performance. For example, how the prediction question is posed (Weaver & Bryant, 1995), how test questions are asked (Maki, Foley, Kajer, Thompson, & Willert, 1990), and how the test question is scored (Maki et al., 1990; Weaver, 1990) are variables that have been shown to affect metacomprehension accuracy. In addition, research suggests that manipulations requiring generative encoding enhance metacomprehension accuracy (Maki et al., 1990). Specifically, Maki et al. found that predictions for paragraphs with deleted letters were more accurate than those for intact paragraphs. According to Maki et al., the necessity of thinking about what the incomplete words should be probably allowed participants to "assess their knowledge" of the material in each paragraph more accurately than simply reading the paragraph would. Similarly, research has demonstrated that metacomprehension improved after participants summarized text material and when participants generated keywords after reading text material (Thiede & Anderson, 2003; Thiede, Anderson, & Therriault, 2003).

The recent studies demonstrating good metacomprehension abilities share one factor. Increasing processing at encoding through generative activities (generation of keywords, Thiede et al., 2003; generating summaries of text material, Thiede & Anderson, 2003; reinserting deleted letters into text, Maki et al. 1990) improves people's ability to predict how they will perform on a later test of memory. These metacomprehension findings are consistent with memory research that suggests that increasing generative activity at encoding improves long-term retention (McDaniel & Einstein, 2005). For example, better retention results from generating answers than from reading answers (McDaniel, Waddill, & Einstein, 1988; McNamara & Healy, 1995). More generally, these findings suggest that difficulty and challenge experienced by the learner may be desirable and should be introduced into instructional practice (e.g., Bjork, 1994; McDaniel & Einstein, 2005). Bjork termed manipulations that require learners to generate during encoding *desirable difficulties*. While researchers have demonstrated that such desirable-difficulty manipulations will positively benefit retention (McDaniel et al., 1988; McNamara & Healy, 1995), recent metacomprehension research mentioned above suggests that desirable difficulty may also play a critical role in accuracy of predictions of test performance. A central objective of the present study is to more fully explore the implications of desirable difficulties on metacomprehension accuracy.

Desirable Difficulty and Retention

One limitation to specifying potential desirable difficulties in terms of the encoding situation (i.e., generative encoding activities) is that this focus does not address processes required by the testing situation. The need to consider the relationship between encoding and test situations for gaining leverage on desirable difficulty was emphasized by Morris et al. (1977), who stated "the value of a particular acquisition activity must be defined relative to particular goals and purposes" (p. 528). Aside from the theoretical importance of including TAP, there is great and underappreciated practical importance. Consider the example related to us by a colleague (McDaniel, 2007) of the high school history instructor who required his students to engage in an integrative activity that fostered analysis of political, economic, religious, and scientific dimensions across cultures. To assess learning, the instructor typically administered a multiple-choice test of cultural details. According to the TAP framework, students in this class may not profit in terms of test performance from this (presumably desirable) study procedure, because the conceptual processes instantiated during acquisition may not correspond to those instantiated by a multiple-choice test of historical details.

To test this general expectation, in the present study we manipulated both the encoding and retrieval situation in order to elicit congruency or incongruency between processes required by study and those required by test. McDaniel, Hines, Waddill, and Einstein (1994, Experiment 1) demonstrated that performance on a cued recall test that relied on detailed information was improved when the study task focused on details (letter reinsertion) rather than on relations among ideas (sentence sorting). For a cued recall task that relied on higher order conceptual information, the opposite pattern was obtained: The study activity that focused on relations among ideas produced better memory performance than the one that focused on details. McDaniel et al. (1994) therefore concluded that the letter reinsertion task facilitated item-specific processing at encoding which, in turn, yielded improved performance on an item-specific test of memory rather than on a relational test of memory. The opposite conclusion was drawn for the sentence-sorting generative task. On the basis of TAP and McDaniel et al. (1994), we expected that the extent to which the generative encoding conditions (letter reinsertion, sentence sorting) reflect desirable difficulty (for learning) would hinge on the congruency between the processing stimulated by the generative task and the processing demands required by the retrieval task. In this study we also examined the more provocative possibility that incongruency between generative encoding and retrieval task could produce memory decrements relative to a nongenerative control.

Desirable Difficulty and Metacomprehension

A second major question of interest is whether TAP is an important factor in the desirability of difficulty for influencing metacomprehension accuracy. The principles of TAP have been applied to the metamemory literature as transfer-appropriate monitoring (Dunlosky & Nelson, 1997). According to the transfer-appropriate monitoring hypothesis, the accuracy of people's predictions of future memory performance is a function of the match between the properties of the judgment context and the properties of the test context (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Dunlosky & Nelson, 1997; Maki & Serra, 1992; Weaver & Kelemen, 2003). Transfer-appropriate monitoring has been empirically tested in research focusing on the accuracy of people's judgments of learning for simple materials such as paired associates, letter strings, and individual words (Koriat, 1997; Weaver & Kelemen, 2003). Integral to the transfer-appropriate monitoring account is the context of the judgment as it relates to the context of the test. However, what evaluations of TAM have not examined is whether prediction accuracy will be affected by the congruency between the context of encoding (rather than judgment) and that of the test.

In order to examine whether transfer-appropriate considerations affect metacomprehension accuracy, the present study used encoding manipulations previously shown to instantiate specific processing of material; that is, generative study activities previously shown to instantiate relational (sentence sorting) or item-specific (letter reinsertion) processing were paired with conceptual or detail-oriented tests. This is the first study to examine whether congruency between processes instantiated by the generative study activity and those required by the test will improve metacomprehension accuracy. To further broaden the implications of the study, we used educationally relevant material.

The Present Study

Processing at encoding was manipulated through generative encoding manipulations. Participants engaged in either a letter reinsertion task or a sentence sorting task, or they simply read the text. Processing at retrieval was manipulated through the type of test questions asked. Participants answered either conceptual or detailed test questions. Processing between encoding and retrieval were congruent when participants engaged in letter reinsertion and were tested with detailed questions and when sentence sorting was paired with conceptual questions. Metacomprehension monitoring was measured in Experiments 1 and 2 by having participants make section-by-section predictions about future test performance. Methodologically, this is a novel way to investigate metacomprehension accuracy, because only one prediction per text passage is usually made, thus limiting the sensitivity of participants' abilities to anticipate future test questions.

Several competing predictions regarding metacomprehension can be derived. First, perhaps previous studies are limited to the generative activities or texts used. Difficulty may therefore not be generally desirable for metacomprehension accuracy. If so, the generative study activities used in the present study would have no effect on metacomprehension accuracy. A second prediction is that any increase in processing at encoding will yield more accurate metacomprehension predictions, regardless of the test context (Maki et al., 1990). Finally, if metacomprehension accuracy is modulated by transfer-appropriate considerations of desirable difficulty, generative processes appropriate to test questions should yield more accurate predictions than would either inappropriate generative processes or a reading study condition.

In addition to investigating prediction accuracy for text material, the present study examines the relationship between those predictions and subsequent controlled processing. In Experiment 2, participants were allowed to restudy sections of previously encoded paragraphs. We wanted to know whether or not the appropriateness of the generative encoding tasks as initially assigned would affect the subsequent study strategies that participants adopted. Even if the kinds of strategies were not affected, subsequent study strategies could be less effective if metacomprehension predictions were rendered less accurate in the inappropriate study–test conditions because the strategy would be based on less-accurate information about anticipated memory performance.

EXPERIMENT 1

Method

Subject and Design. Thirty-nine Washington University undergraduate students received course credit for their participation. The average age of participants was 19.60; their average years of education were 14. The design was a 3 (type of encoding task: letter reinsertion, sentence sorting, reading) $\times$ 2 (type of test: detailed vs. conceptual) within-subjects factorial design.

Materials. The expository texts used were "Kanchenjunga: A Very Dangerous Mountain Range," "The Frozen Continent," "How Autumn Colors are Formed," "Nomads of the Desert," "The Strange Way of Spiders," and "Skunks" (taken from Levy, 1981). The average word count of these passages was 326; they were chosen so that participants would be able to make six predictions per passage, and so that six cued recall questions could be answered at the end. For each passage, six detail-oriented questions and six conceptual questions were generated. The detailed questions assessed individual item information and the conceptual questions assessed global thematic information. An example of a detailed question is "The walls of ice in Kanchenjunga range from _______________ feet high." The answer is "600 to 1,000." An example of a thematic question is "The downward speed of snow is much faster in Kanchenjunga than in the Alps because _________________________." The answer is "there is more snowfall." The thematic questions related information that was presented across multiple sentences. Depending on condition, for a given passage participants received either detail-oriented questions or conceptual questions. Each passage was divided into six sections with approximately equal numbers of words in each section.

Four practice passages were also used in this experiment. They were expository in nature and, averaging 181 words each, shorter than those used throughout the rest of the experiment. Although they were shorter, practice passages were matched on reading comprehension with experimental passages. Reading comprehension was determined by having a separate group of 20 participants read and answer detailed and conceptual questions associated with practice and experimental passages. Scores on this test determined whether the passage would be selected for the study. In order for a passage to be selected for the study, the average percentage of correct answers on the associated test had to be between 80 and 85. The practice passages used were "Sharks of Hawaii" (State of Hawaii, Division of Aquatic Resources, Department of Land and Natural Resources, 2004); "English Bulldogs" (english-bulldogs.org, 2004), "Leonardo da Vinci, Renaissance Man" (Museum of Science, Boston, 2004), and "Alaskan Rain Forest: The Land and Its People" (Alaska Rainforest Campaign, 2004). Practice passages were matched with experimental passages on level of comprehension. Practice passages were given to orient participants to the encoding task (letter reinsertion vs. sentence sorting) and type of test (detailed vs. conceptual).

Procedure. As an overview, the procedure included a practice portion with four expository texts (two encoded with sentence sorting and two with letter reinsertion) followed by presentation of the six target passages (two sentence sorting, two letter reinsertion, and two read-only). For each passage, participants performed a given encoding/orienting task, completed metacomprehension predictions, and performed a cued-recall task. Tasks were blocked for a given passage so that the passage would be encoded, predictions made, and the test completed before presentation of the next passage.

For the practice phase, participants were told that passages would be presented and that they would have one of two tasks: to fill in letters, or to reorder sentences. The orienting task was randomly determined and participants were alerted as to what the orienting task was going to be when the title of the passage appeared on the computer screen along with the orienting task that was to be performed on the passage. The experimenter gave participants a paper copy of the passage with letters missing. Participants were then told to fill in the letters in the passage to make the passage coherent; they were further instructed to complete the task as quickly and accurately as possible. The passage was presented in a normal sequence, but approximately 15% of the letters, of which approximately 40% were vowels, were deleted. The specific locations of the letter reinsertions were haphazardly determined and several adjustments were made in pilot research so that most participants could identify all of the words. For sentence reordering, participants were given their passages on slips of paper fastened by Velcro to larger sheets of paper. Sentence positions were set at no more than 4.0 positions away from the original, correct position in the unscrambled version of the passage. In addition, two sentences were always yoked together. The passage was reordered in a predetermined fashion and the task of the participants was to correctly order the passage.

After the encoding task, participants were presented with the passage again so that metacomprehension predictions could be made. Because we were interested in whether knowledge of the type of test influenced metacomprehension predictions, participants were told what type of test to expect for each passage after the orienting task but before the metacomprehension predictions were made. More specifically, participants were presented again with the title of the passage that had previously been encoded and they were told that they would be taking a test that either focused on details from the passage or on conceptual themes that connected concepts within the passage. Participants were told to use the information about the upcoming test to guide them when making metacomprehension predictions. The passage was then presented again, with all letters intact and in the correct order. The passages were divided into six sections of two to four sentences and presented on IBM-compatible computers. Presentation of passage sections was self paced; participants pressed the space bar when they were ready to make the metacomprehension prediction associated with the previously presented section. The question on metacomprehension prediction was "How well do you think you will remember the information in the previous sentences on an upcoming test?" The prediction was made on a scale where 0% represented "not likely to remember at all" and 100% represented "extremely likely to remember." After all six sections of the passage had been presented and rated, participants were asked to make a global metacomprehension judgment; specifically, they were asked "How likely are you to remember the information contained in the previous passage?" This prediction was made on a Likert scale, where 0 represented "not likely to remember at all" and 10 represented "extremely likely to remember."

After making the final global prediction, participants had to complete a detailed or conceptually oriented cued recall test for each passage. Questions were presented on the computer and the test was self paced. Participants read questions and entered their answers, which had to be of 10 words or fewer. Participants were allowed to type "Do not know" if they could not answer a question. Once an answer was given, participants pressed the Enter key and the next question was presented. As mentioned earlier, six questions were presented for every passage read.

Once those three components were completed for a given passage, the next practice passage was presented. After participants practiced on the four practice texts, they began the experimental session. They were told that they would be presented with six passages. Two passages would be encoded through sentence sorting, two through letter reinsertion, and two would simply be read. The experimental phase followed the practice procedure exactly. Material assignment and order of encoding condition were randomly determined; that is, a computer program was designed to select a passage at random and to assign that passage to a given encoding condition. This selection was random but without replacement, so that all passages were encoded, and every encoding condition was employed twice for a given participant. The read condition was included to examine metacomprehension accuracy for each type of test (detailed or conceptual) in the absence of explicit encoding manipulations.

Table 1
Mean Cued Recall Performance As a Function of Orienting Task and Type of Test

	Detailed		Conceptual	
	M	*SE*	*M*	*SE*
Experiment 1				
Letter reinsertion	.60	.04	.30	.03
Sentence sorting	.43	.04	.63	.04
Reading	.43	.05	.49	.06
Experiment 2				
Letter reinsertion	.84	.03	.25	.04
Sentence sorting	.37	.07	.85	.04

Results

Cued recall performance. Table 1 displays mean cued recall performance as a function of type of task and type of test. A 3 (type of task: letter reinsertion, sentence sorting, reading) × 2 (type of test: detailed, conceptual) within-subjects ANOVA on the proportion correctly recalled confirmed that the interaction between type of task and type of test was significant [$F(2,36) = 10.34$, $MS_e =$ .06]; the rejection level for statistical significance was set at .05, unless otherwise stated. As can be seen in Table 1, when the encoding task stimulated focus on information presumably required for the test, performance improved (relative to reading). However, when the encoding task did not force processing of the information needed for the test, no improvement occurred. A main effect for type of task was also found [$F(2,36) = 3.26$, $MS_e = .05$]. Participants who engaged in sentence sorting at encoding ($M =$ .55) generally performed better on the cued recall tasks than did participants who engaged in letter reinsertion ($M = .45$); [$t(26) = 3.13$, $SE = .04$]. While a nominal difference was apparent between sentence sorting and reading, the difference was not statistically significant. No other effects were significant.

Mean metacomprehension predictions. Mean metacomprehension predictions as a function of type of orienting task and type of test are presented in Table 2. A 3 (type of task: letter reinsertion, sentence sorting, reading) × 2 (type of test: detailed, conceptual) ANOVA performed on mean predictions found a significant interaction between type of task and type of test [$F(2,48) = 10.03$, $MS_e = .01$]. As can be seen in the top section of Table 2, the letter reinsertion task elicited lower prediction ratings when a conceptual test was expected than it did when a detailed test was expected [$t(33) = 3.73$, $SE = .03$]. The sentence sorting task elicited higher overall predictions when participants expected the conceptual test than when participants expected a test that tapped details [$t(38) =$ 2.67, $SE = .03$]. For the read condition, there was a slight

Table 2
Mean Metacomprehension Predictions As a Function of Orienting Task and Type of Test

	Detailed		Conceptual	
	M	*SE*	*M*	*SE*
Experiment 1				
Letter reinsertion	.62	.02	.53	.03
Sentence sorting	.55	.03	.63	.02
Reading	.55	.03	.59	.03
Experiment 2				
Letter reinsertion	.64	.03	.50	.03
Sentence sorting	.53	.03	.61	.03

and nonsignificant higher prediction when participants were given conceptual test questions ($t < 1$). The average global metacomprehension rating was 5.3, and no significant differences were found when global metacomprehension judgments were subjected to a similar analysis.

Accuracy of ratings. As in previous studies (e.g., Glenberg et al., 1987; Maki & Serra, 1992; Thiede et al., 2003; Weaver, 1990), metacomprehension accuracy was operationalized as a Goodman–Kruskal gamma correlation between a participant's prediction and actual test performance. For each participant, a gamma correlation was computed. The mean of these intraindividual correlations was then computed across participants within each group. This method of analysis has been shown to not be affected by an individual's level of test performance or absolute threshold of prediction (for further discussion, see Nelson, 1984). The means of these intraindividual correlations are displayed in Table 3 as a function of condition, and, as can be seen, both orienting task and type of test affected metacomprehension accuracy. These data were subjected to a 3 (type of task: letter reinsertion, sentence sorting, reading) × 2 (type of test: detailed, conceptual) ANOVA. A main effect of type of task was found [$F(2,76) = 4.46$, $MS_e = .23$]. Pairwise comparisons indicate that sentence sorting (collapsed across type of test) led to significantly more accurate predictions ($M = .26$) than occurred with letter reinsertion ($M = .04$) [$t(38) = 2.57$, $SE = .09$]. The read condition ($M = .19$) (also collapsed across test) led to significantly more accurate predictions than did the letter reinsertion condition [$t(38) = 1.80$, $SE = .08$]. There was no difference in prediction accuracy between the sentence sorting condition and the read condition.

Most importantly, an interaction between type of test and type of task was found [$F(2,76) = 9.55$, $MS_e = .17$]. To examine the nature of this interaction, planned comparisons were performed to examine how each study task influenced metacomprehension accuracy for each test, relative to the read conditions. For the sentence sorting task, when participants expected a conceptual test they made more accurate predictions ($M = .59$) than when they expected the read–conceptual condition ($M = .11$) [$t(38) = 6.56$, $SE = .07$]. In contrast, after participants completed the sentence sorting task, and when they expected detailed test questions, they made less accurate predictions ($M = -.06$) than they did for the read–detailed condition ($M = .26$) [$t(38) = 3.54$, $SE = .09$]. For the letter reinsertion task there was a significant impairment in metacomprehension accuracy when a conceptual test was expected ($M = -.33$) relative to a read–conceptual condition [$t(38) = 4.21$, $SE = .10$]. The differences between letter reinsertion–detailed and read–detailed did not yield a significant difference ($t < 1$). Thus, when the study task was incongruent with the test (letter reinsertion–conceptual; sentence sorting–detailed), prediction accuracy was impaired, compared with the read–conceptual condition. When the two were congruent, metacomprehension accuracy improved over that of the read condition (for the sentence sorting–conceptual condition).

Finally, we examined whether participants in each condition produced gamma correlations that were significantly above (or below) zero. The following conditions yielded mean gamma correlations significantly higher than zero: letter reinsertion–detailed [$t(38) = 5.02$, $SE = .08$]; sentence sorting–conceptual [$t(38) = 8.50$, $SE = .07$]; read–detailed [$t(38) = 4.26$, $SE = .06$]; read–conceptual [$t(38) = 2.10$, $SE = .05$]. In contrast, the sentence sorting–detailed condition did not yield correlations significantly different from zero, and the letter reinsertion–conceptual condition produced correlations significantly less than zero [$t(38) = -3.75$, $SE = .08$]. A negative correlation would suggest that the incongruency between encoding and retrieval tasks led to complete miscalibration, whereby participants thought they had mastered material when they had not. This negative correlation would also suggest that participants thought they did not have mastery of material when they did. However, because this miscalibration was not seen in both incongruent conditions, it seems unlikely that the finding generalizes across a range of incongruent task–test circumstances.

Discussion

The first salient finding was that cued recall of text information was enhanced when the type of generative study activity focused on information congruent with that required in the cued recall test. This pattern replicates that reported by McDaniel et al. (1994, Experiment 1). The results in the present study extend that initial finding with more educationally representative texts. More generally, these results are in line with the basic transfer-appropriate processing literature (using typical laboratory stimuli such as word lists), showing that memory performance is enhanced when the orienting activities stimulate

Table 3
Mean Gamma Correlations Between Predicted and Actual Performance As a Function of Orienting Task and Type of Test

	Detailed		Conceptual	
	M	*SE*	*M*	*SE*
Experiment 1				
Letter reinsertion	.41	.08	−.33	.08
Sentence sorting	−.06	.07	.59	.07
Reading	.27	.06	.11	.05
Experiment 2				
Letter reinsertion	.62	.05	−.13	.13
Sentence sorting	.06	.15	.55	.07

processing of information required in the test task (e.g., Blaxton, 1989; McDaniel et al., 1978; Morris et al., 1977; Nairne & Widner, 1987). These results suggest that there are limitations to the use of difficulty manipulations at encoding. Difficulty manipulations yield improvement in performance, beyond that of simply reading texts, only when the testing situation requires processing congruent with that instantiated by the generative study task. In fact, when the processing required by the study task was incongruent with processing required by the test, cued recall performance was worse than when participants were not given a generative study task and were simply instructed to read the passage instead. Thus, these results underscore the importance in educational settings of assigning study activities that produce processing congruent with the information being emphasized during testing. Indeed, the present experiment demonstrates that encoding processes that accentuate processing of details produce a reduction in cued recall for conceptual test items relative to the read-only condition. These findings imply that assigning study activities incongruent with the test task can have a negative impact on performance.

Of additional importance are the metacomprehension findings. Similar to the findings of Weaver and Bryant (1995), when participants were given expository texts to read, prediction accuracy was greater when detailed test questions were given than when conceptual test questions were given. Additionally, as with previous research, when participants were given the opportunity to make multiple metacomprehension predictions (Weaver & Bryant, 1995) and were given multiple test questions for a given passage (Maki et al., 1990), they were able to accurately predict how much information they would later be able to recall.

More novel are the findings showing how study activities affect metacomprehension accuracy. First, metacomprehension improved relative to reading, when type of generative study task and type of test were congruent. More startling was the finding that metacomprehension was completely disrupted when type of generative study task was incongruent with type of test (i.e., letter reinsertion–conceptual test; sentence sorting–detailed test). In these conditions, prediction accuracy fell to chance or worse. Indeed, in the letter reinsertion–conceptual condition, we found that participants were significantly miscalibrated when making predictions.

Two aspects of these metacomprehension failures are notable. First, they were not tied specifically to reductions in memory. The sentence sorting condition produced detail cued recall levels identical to that of the read condition, yet metacomprehension accuracy was significantly reduced to levels below zero (relative to read). Second, metacomprehension impairment in the incongruent study and test conditions was not a consequence of inappropriate modulation of absolute prediction levels; that is, the mean prediction level associated with detailed tests did not vary across the sentence sorting and read conditions (where recall levels remained the same). Similarly, the mean prediction level for conceptual cued recall was lower in the letter reinsertion conditions than in the read conditions, paralleling the actual recall performances in these conditions. Given the observations above, a plausible interpretation is that on a more global level, participants can modulate predictions based on both congruent and incongruent study–test processing. That is, on average, participants may be aware that they will be able to recall less information in incongruent versus congruent conditions; however, when asked to make fine-grained section-by-section judgments, participants are unable to predict performance in incongruent conditions. On the other hand, when processing is congruent, metacomprehension accuracy improves relative to the read conditions, suggesting that participants are aware of section mastery and aware of failure to master. Thus, test-appropriate study activities confer compounded benefits: They increase recall for appropriate tests, and they increase metacomprehension accuracy.

These findings lead to the provocative possibility that at least a third positive benefit might accrue to test-appropriate study activities, a benefit that has significant educational implications. Specifically, given the high level of metacomprehension accuracy for test-appropriate study tasks, these tasks may promote efficacious control of subsequent learner-initiated study activities. In contrast, the decrement in metacomprehension accuracy when test-inappropriate study activities were used may extend to prompt ineffectual and misguided control of subsequent self-study. Experiment 2 was conducted to test these possibilities, as well as to replicate the novel metacomprehension findings obtained in Experiment 1.

EXPERIMENT 2

Like Experiment 1, Experiment 2 compared metacomprehension prediction accuracy for congruency and incongruency in encoding and retrieval processes. We expected to replicate the findings of Experiment 1, which suggested that people have difficulty predicting future test performance when the type of task is incongruent with test questions. According to the *monitoring affects control* hypothesis (Nelson & Leonesio, 1988), controlled processes such as study-time allocation are dependent on monitoring—metacomprehension predictions, for example. If predictions are not correlated with later test performance, it is unlikely that learners will implement effective controlled processing to relearn material. Consequently, it may matter little what study-time allocation policy learners adopt in this situation, as the items selected for study will have no bearing on their actual difficulty. However, in the congruent encoding activity–test conditions, Experiment 1 showed that people can accurately predict future performance. In this situation, the study-time allocation policies should be more crucial for determining memory performance because the learners' judgments of learning will be accurate. What remains unknown is the nature of the allocation policy adopted by participants.

Early research examining self-paced study demonstrated that people do allocate different amounts of study time to various items in paired associates learning (Zacks, 1969), in serial recall (Le Ny, 1969), and in free recall (Belmont & Butterfield, 1971). The subsequent research examining study-time allocation converges on the conclusion that people tend to allocate their study time to items they judge to be difficult (Cull & Zechmeister, 1994; Dunlosky & Connor, 1997; Dunlosky & Hertzog, 1997; Kellas & Butterfield, 1971; Le Ny, Denhiere, & Le Taillanter, 1972; Mazzoni & Cornoldi, 1993; Mazzoni, Cornoldi, & Marchitelli, 1990; Mazzoni, Cornoldi, Tomat, & Vecchi, 1997; Nelson, Dunlosky, Graf, & Narens, 1994; Nelson & Leonesio, 1988; Thiede & Dunlosky, 1999). The present experiment investigated whether participants would allocate more study time to items judged to be more difficult in both congruent and incongruent task–test processing conditions. Further, this experiment investigated whether the allocation of additional study time would lead to an improvement in overall cued recall performance.

Method

Subjects and Design. Twenty-five Colby College undergraduate students received course credit for their participation. The average age of participants was 18.70, and their average years of education were 13. The design was a 2 (type of encoding task: letter reinsertion vs. sentence sorting) × 2 (type of test: detailed vs. conceptual) within-subjects factorial design. The read condition was not used in this experiment because of the null effect in both cued recall performance and metacomprehension accuracy when type of test was compared.

Materials and Procedure. The materials were those used in Experiment 1. The procedure was similar to that of Experiment 1, except for the addition of one stage. Specifically, participants first engaged in an encoding phase during which they were given expository texts and had to either perform letter reinsertion or sentence sorting; each study activity was therefore performed on three target texts. After performing the orienting task, participants made metacomprehension predictions. The predictions were collected in the same fashion as in Experiment 1. After predictions, for further study, participants were presented again with the same sections of the passage as those used when metacomprehension predictions were collected. Before the presentation of the first section of a given passage, participants were told that they could spend as much or as little time restudying the section as they deemed necessary. Because a given passage had already been processed two times (once during the initial encoding phase, and a second time during the metacomprehension prediction phase), participants could, if they chose, skip rereading a given section of a passage by simply pressing the space bar. On the other hand, participants could spend as much time as they chose trying to master the material in a given section of the passage.

After the restudy phase, participants were presented with the cued recall test. These tests were identical to those used in Experiment 1. Further, as in Experiment 1, participants first engaged in a practice phase with four practice passages. The practice phase was identical to that implemented in Experiment 1, with the addition of the restudy stage.

Results

Cued recall performance. The bottom section of Table 1 illustrates the average cued recall findings for this experiment. To analyze these results, a 2 (type of task: letter reinsertion, sentence sorting,) × 2 (type of test: detailed, conceptual) within-subjects ANOVA was performed on the proportion correct on the cued recall test. While neither the main effect for type of study task nor type of test was significant ($F < 1$), the interaction between the two was significant [$F(1,24) = 100.28$, $MS_e =$ .07]. As in Experiment 1, when the letter reinsertion task was paired with detailed test questions, cued recall performance was better than when letter reinsertion was paired with conceptual test questions. Similarly, when sentence sorting was paired with conceptual test questions, cued recall performance was better than when sentence sorting was paired with detailed test questions.

To investigate the effects of restudy, we examined the results of Experiments 1 and 2 in a 2 (type of task: sentence sorting, letter reinsertion) × 2 (type of test: detailed, conceptual) × 2 (type of study: restudy, no restudy) mixed ANOVA, with type of task and type of test as the within-subjects variables and type of study as the between-subjects variable. Cued recall performance increased with the opportunity for restudy from $M = .49$ to $M = .58$ [$F(1,53) = 3.5$, $MS_e = .08$]. Importantly, the three-way interaction showed that the improvement in performance related to restudy was limited to the congruent processing conditions [$F(1,53) = 19.49$, $MS_e = .06$] (see Table 1).

Mean metacomprehension ratings. Mean metacomprehension predictions as a function of type of orienting task and type of test are presented in the bottom section of Table 2. A 2 (type of task: letter reinsertion, sentence sorting) × 2 (type of test: detailed, conceptual) ANOVA performed on mean metacomprehension predictions found a significant interaction between type of task and type of test [$F(1,24) = 6.58$, $MS_e = .02$]. These results parallel those found in Experiment 1. The average global metacomprehension prediction was 5.8 and no significant differences were found when judgments were subjected to a similar analysis.

Accuracy of ratings. As can be seen in the bottom section of Table 3, the results of Experiment 2 replicated those found in Experiment 1, in that a match between encoding task and type of test led to more accurate predictions than when encoding task and test were incongruent. To analyze these findings, mean gamma correlations were subjected to a 2 (type of task: letter reinsertion, sentence sorting) × 2 (type of test: detailed, conceptual) ANOVA. The crossover interaction between type of test and type of task was significant [$F(1,24) = 6.32$, $MS_e = .30$]. When type of task and type of test were congruent, metacomprehension predictions were more accurate than when the two were incongruent. In addition, we examined whether participants in each condition produced gamma correlations that were significantly above (or below) zero. Only the congruent conditions yielded mean gamma correlations that were significantly different from zero: letter reinsertion–detailed [$t(24) = 9.24$, $SE = .06$]; sentence sorting–conceptual, [$t(24) = 6.78$, $SE = .07$]. In contrast, neither of the incongruent conditions (letter reinsertion–conceptual, sentence sorting–detailed) led to gamma correlations that were significantly different than zero.

Study-time allocation on the basis of predictions. Recall that participants had the opportunity to restudy any section of the previously presented passages after metacomprehension predictions had been made. As with metacomprehension predictions, participants were presented again with ordered segments of each passage. Each participant could spend as much or as little time on each section as deemed necessary. Study-time allocated to a given section was then correlated with predictions of future performance. Figure 1 presents the average gamma correlations. These data were subjected to a 2 (type of task: letter reinsertion, sentence sorting) × 2 (type of test: detailed, conceptual) within-subjects ANOVA. There were no significant main effects or interactions ($F < 1$). Consistent with the *monitoring effects control hypothesis* (Nelson & Leonesio, 1988), participants allocated study time on the basis of their metacomprehension prediction. More time was allocated to items judged to be less likely to be recalled and less time to items judged to be more likely to be recalled. This allocation policy was not affected by study task or type of test. However, it is important to note that the four correlations (between study-time allocation and metacomprehension predictions) were significantly different from 0 [letter reinsertion–detailed: $t(19) = -2.3$, $SE = .2$; letter reinsertion–conceptual: $t(22) = -3.8$, $SE = .3$; sentence sorting–detailed: $t(23) = -2.8$, $SE = .2$; sentence sorting–conceptual: $t(21) = -3.5$, $SE = .3$]. While monitoring had a significant impact on control, because it was impaired in the incongruent conditions, the controlled processes implemented were ineffective, as evidenced by cued recall performance.

Discussion

Just as in Experiment 1, we found that cued recall performance was better in conditions where processes instantiated by encoding were congruent with those at test; that is, when letter reinsertion was paired with detailed test questions, participants performed better than when letter reinsertion was paired with conceptual test questions. These results are consistent with the TAP framework (Blaxton, 1989; McDaniel et al., 1978; Morris et al., 1977, Nairne & Widner, 1987). More novel was the finding that restudy had little benefit for retrieval when processes at encoding and retrieval were incongruent, but cued recall performance significantly improved from Experiment 1 to Experiment 2 in congruent conditions. For example, the gain from Experiment 1 to Experiment 2 for sentence sorting–conceptual was .21; the gain for sentence sorting–detailed was −0.3. A parallel pattern was present for the letter insertion conditions. These results suggest that participants could select the appropriate items for restudy in congruent conditions. By contrast, because monitoring was impaired in incongruent conditions, participants were unable to effectively select items for restudy in these conditions.[1] We interpret these results as indicating that participants were more successful in selecting appropriate controlled strategies in congruent conditions, but it is also possible that retention improvements found in Experiment 2 might have been caused by overall differences in study time in the congruent conditions. However, average study time did not differ as a function of generative study task or type of test.

Experiment 2 also clearly replicated the correlational findings between predictions and performance found in Experiment 1; that is, people were better calibrated in congruent than in incongruent conditions. The important contribution of Experiment 2 is how monitoring as assessed by metacomprehension predictions affected later cognitive resource allocation. Control in this experiment was measured by restudy time. When restudy time was correlated with metacomprehension predictions, we found that participants selected items to restudy that were judged to be less likely to be recalled. These findings are consistent with much of the study-time allocation literature, which has found that learners allocate more study time to items judged difficult (Son & Metcalfe, 2000). Although these findings are consistent with the literature, their implications are somewhat troublesome. Specifically, in the incongruent condition, participants allocated time to items judged harder; however, we know from the correlational analysis between predictions and performance (Experiments 1 and 2) that items judged more difficult were not necessarily so. Thus, in the incongruent conditions, though monitoring may affect control (allocation of study to harder items), monitoring was inaccurate and the resulting controlled process may accordingly have been ineffective.

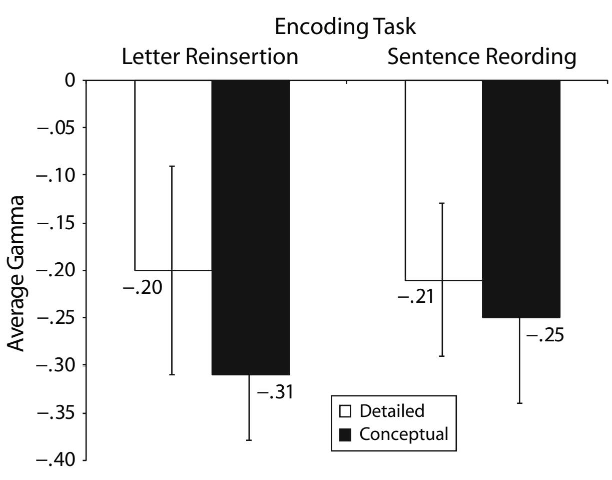

Figure 1. Experiment 2 gamma correlations between metacomprehension predictions and study-time allocation.

GENERAL DISCUSSION

The primary motivation for these experiments was to investigate whether or not generative study activities would improve retention and metacomprehension. Previous research has found that generative study activity does improve both retention (McDaniel et al., 1988; McNamara & Healy, 1995) and metacomprehension (Maki et al., 1990), though few studies have examined the interaction between generative study manipulations and testing on retention and metacomprehension of educationally relevant material. As an overview, we found that when processing at encoding and test were congruent, cued recall and prediction accuracy improved beyond simply reading text material (Experiment 1). Further, we found a significant relationship between study-time allocation and predictions of performance. On the other hand, when processes instantiated by study tasks were incongruent with the information required at retrieval, cued recall was worse than when participants simply read the text and prediction accuracy was at chance or worse. This latter result reveals that participants' intended strategy of allocating more time to difficult items (Experiment 2) could not have been effectively implemented in incongruent study–test conditions. We next elaborate on these findings.

Cued Recall

Previous research suggests that encoding manipulations designed to promote difficulty may enhance retention (Bjork, 1994). However, the present study demonstrates that difficulty at encoding may not always be desirable. The present study tested whether the TAP framework might better assess the utility of generative encoding manipulations (McDaniel et al., 1988). According to TAP, manipulations that encourage a particular type of processing should improve memory when the processing invited by the test is congruent with that of study. We used the letter reinsertion task to invite individual item processing of text, and the sentence sorting task was used to invite relational processing of text (McDaniel et al., 1994; McDaniel et al., 1988). Performance on a cued recall test of details was enhanced by individual item processes, whereas performance on a cued recall test of conceptual information was enhanced by relational processing. These findings suggest that designation of a generative encoding task as desirable may be dependent on how learned information is tested.

More generally, our findings argue for a contextualistic framework for determining desirability of difficulty (McDaniel & Einstein, 2005). In addition, our findings not only support the TAP framework for educationally relevant materials, they also demonstrate that when processes at encoding and retrieval do not match, performance on the criterial task is significantly impaired, compared with mere reading texts. For example, when participants performed letter reinsertion and were given a test of conceptual information, they performed worse than when participants read the text and were given the same test (Experiment 1). For educational settings, the implications are clear. Instructors would be ill advised to require study activities that encourage processing that does not match that needed by the test; yet in some cases we have known of instructors who required students to perform study activities focusing on higher order relations among constructs, and who then administered tests focusing on detailed factual information. This mismatched study activity, though apparently desirable, would be expected to impair test performance relative to allowing students to simply read the material.

Relative and Absolute Accuracy

More novel is our finding that metacomprehension prediction accuracy was also affected by the congruency of encoding and retrieval processes. According to Maki et al. (1990), generative study manipulations probably allowed participants to "assess their knowledge" of the material in each paragraph more accurately than did simply reading the paragraph. Consistent with this hypothesis, Maki et al. demonstrated that more-accurate predictions of performance resulted when participants engaged in increased processing at encoding. While Maki et al. concluded that more processing leads to better assessment of later recallability, the findings from the present study suggest that the relationship between processing and predictions is more complex.

As is evidenced by both experiments, when processes invited by the difficulty manipulation were congruent with those required by the test, the correlation between predictions and performance, or relative accuracy, was greater than when processing was incongruent. Further, Experiment 1 demonstrates that prediction accuracy was greater in congruent conditions than in conditions in which participants did not engage in generative encoding (i.e., read). These results are consistent with those of Maki et al. (1990); however, the present findings refine Maki et al.'s original conceptualization, which stated that *any* additional processing at encoding would lead to improvements in prediction accuracy. Our study demonstrates that when task–test processing is incongruent, prediction accuracy is at chance or worse. Early work examining the effects of processing on metacomprehension had not examined the interaction between the processing invited by study tasks and that required by the test. The present study demonstrates that this interaction is particularly important, and further demonstrates that not all generative study processes will benefit later prediction accuracy. Rather, only when study task–test processing is congruent will we see improvements in prediction accuracy. We assume that the results found by Maki et al. were based on a congruency between processing task and test but they did not describe what type of test questions were used.

Not only do these findings contribute to our understanding of metacomprehension, they have significant applicability to the classroom. In both experiments, we demonstrate conditions in which test performance and prediction accuracy can be enhanced. Elaboration at study will enhance test performance and prediction accuracy only when that elaboration is consistent with how learners will be tested. Thus, creating desirable difficulty (Bjork, 1994) will be beneficial to learners only when the processes invited by those difficulty manipulations are congruent with those required by test. This is the first study to demonstrate that congruency of these two processes will improve both metacomprehension accuracy and later recall. Based on our findings, in congruent study task–test conditions, students should be able to accurately assess their knowledge and therefore study effectively.

While relative accuracy was significantly impaired in incongruent processing conditions, absolute accuracy was not. In both experiments, mean metacomprehension predictions were lower in the incongruent conditions than in the congruent conditions. These results suggest that participants are aware of the processes invited by the specific difficulty manipulations and are also aware of how those processes might interact with those required by the specific tests. (Recall that participants are told what type of test to expect before making metacomprehension predictions.) Participants may be aware on some fundamental level that specific difficulty manipulations may invite specific processes, and that a congruency between study and test processing will lead to improvements in performance; however, they had difficulty applying this knowledge to make more fine-grained, section-by-section predictions. Further support for this conclusion comes from an analysis of the distribution of predictions as a function of type of task and type of test. In both experiments, incongruent conditions led to a lower set midpoint for predictions than did congruent conditions.

Monitoring Affects Control

Incongruent study task–test processing had a negative impact on memory performance. It also rendered metacomprehension predictions useless. If we assume that monitoring affects control, inaccurate metacomprehension predictions should therefore impair one's ability to effectively implement controlled processing. According to the *monitoring affects control* hypothesis (Nelson & Leonesio, 1988), metacomprehension predictions will direct study-time allocation. Interestingly, this hypothesis further indicates that more study time will be allocated to items judged less well learned. In fact, that is exactly what we found when study task–test processing was incongruent. Thus, our data (at least from the incongruent condition) support the monitoring affects control hypothesis; however, our data also call into question the mnemonic effectiveness of this relationship in certain situations. That is, while monitoring may affect control, if monitoring is impaired learners will be unable to implement an effective controlled strategy, and the result will be poor performance on the criterial task. This reflects a negative cascade of incongruent task–test processing.

CONCLUSIONS

These experiments significantly extend the basic metacomprehension literature. Previous work showed that generative encoding activities can produce good metacomprehension for text. The present findings indicate that the metacomprehension benefits of generative encoding activities depend on implementing generative processing congruent with the requirements of the test. Generative study–test congruency is thus a key for the desirable difficulty of generative tasks, with the desirability affecting memory performance, metacomprehension accuracy, and control of further study activities.

The present findings would suggest that generative study–test congruency leads to metacomprehension accuracy, because congruency yields a rich relevant-knowledge base on which to base metacomprehension predictions. While relevant knowledge appears to be instrumental for accuracy in metacomprehension, other factors may also play a significant role. For example, anticipation of test questions when making metacomprehension predictions may affect accuracy of those predictions. Admittedly, having access to relevant knowledge may imply that one is better able to anticipate test questions. One has to be aware that one's knowledge is indeed relevant, but a learner cannot be sure of this without previous experience with similar types of test questions. In the present study, participants practiced with both detailed and conceptual test questions. Participants were also told what type of test questions to expect before making metacomprehension predictions. Familiarity with, and expectations of, test questions may also have affected prediction accuracy. Future research should be directed at examining the contribution of relevant knowledge, test question familiarity, and test type expectation.

From an applied perspective, our results have direct implications for how instructors teach classes, how textbooks are constructed, and how students are trained to study for tests. As we have known for some time, when processing invited by the study activity is congruent with that invited by the test, good test performance will result. The present study extends this finding to more complex material, and further demonstrates that, when study task–test processing is incongruent, performance can fall below levels achieved by learners who only read the material. Thus, if students are instructed to study the details in their textbooks and are given a test that focuses on thematic relationships, they may perform significantly worse on that test than if they had not been instructed at all. Our results have further demonstrated that study task–test incongruency produces a negative cascade of events; memory performance is negatively affected, monitoring falls to chance levels, and intended controlled processing is rendered ineffective. These results suggest that the ability to effectively study relies heavily on establishing congruent relationships between encoding activities and retrieval tasks.

AUTHOR NOTE

This research was supported by a James S. McDonnell Foundation Collaborative Activity Award (Grant #220020041). We thank Luciana Bonifaz and Julie Kling for their help on various aspects of this project, including subject testing and data scoring. We also thank Janet Metcalfe, John Dunlosky, and an anonymous reviewer for commenting on an earlier version of the manuscript. Correspondence concerning this article should be addressed to A. K. Thomas, Department of Psychology, Tufts University, 490 Boston Ave., Medford, MA 02155 (e-mail: ayanna.thomas@tufts.edu).

REFERENCES

ALASKA RAINFOREST CAMPAIGN (2004). *Alaska rainforest: The land and its people.* Retrieved September 15, 2004 from www.akrain.org/rainforest/theland.

BEGG, I., DUFT, S., LALONDE, P., MELNICK, R., & SANVITO, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory & Language*, **28**, 610-632.

BELMONT, J. M., & BUTTERFIELD, E. C. (1971). Learning strategies as determinants of memory deficiencies. *Cognitive Psychology*, **2**, 411-420.

BJORK, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

BLAXTON, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 657-668.

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. *Journals of Gerontology, Series B: Psychological Sciences & Social Sciences*, **52B**, P178-P186.

Dunlosky, J., & Nelson, T. O. (1997). Similarity between the cue for judgments of learning (JOL) and the cue for test is not the primary determinant of JOL accuracy. *Journal of Memory & Language*, **36**, 34-49.

English-bulldogs.org (2004). *English Bulldogs.* Retrieved September 15, 2004, from www.english-bulldogs.org.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, **34**, 906-911.

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 702-718.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, **116**, 119-136.

Glenberg, A. M., Wilkinson, A. C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, **10**, 597-602.

Kellas, G., & Butterfield, E. C. (1971). Effect of response requirement and type of material on acquisition and retention performance in short-term memory. *Journal of Experimental Psychology*, **88**, 50-56.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, **126**, 349-370.

Le Ny, J. F. (1969). Studies of study: The effect of position in intentional learning of a series. *L'Année Psychologique*, **69**, 81-92.

Le Ny, J. F., Denhiere, G., & Le Taillanter, D. (1972). Regulation of study-time and interstimulus similarity in self-paced learning conditions. *Acta Psychologica*, **36**, 280-289.

Levy, B. A. (1981). Interactive processes during reading. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 1-35). Hillsdale, NJ: Erlbaum.

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 663-679.

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 609-616.

Maki, R. H., & Serra, M. (1992). The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 116-126.

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.

Mazzoni, G., Cornoldi, C., Tomat, L., & Vecchi, T. (1997). Remembering the grocery shopping list: A study on metacognitive biases. *Applied Cognitive Psychology*, **11**, 253-267.

McDaniel, M. A. (2007). Rediscovering transfer as a central concept. In H. L. Roediger III, Y. Dudai, & S. Fitzpatrick (Eds.), *Science of memory: Concepts*. New York: Oxford University Press.

McDaniel, M. A., & Einstein, G. O. (2005). Material appropriate difficulty: A framework for determining when difficulty is desirable for improving learning. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications* (pp. 73-85). Washington, D.C.: American Psychological Association.

McDaniel, M. A., Friedman, A., & Bourne, L. E., Jr. (1978). Remembering the levels of information in words. *Memory & Cognition*, **6**, 156-164.

McDaniel, M. A., Hines, R. J., Waddill, P. J., & Einstein, G. O. (1994). What makes folk tales unique: Content familiarity, causal structure, scripts, or superstructures? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 169-184.

McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory & Language*, **27**, 521-536.

McNamara, D. S., & Healy, A. F. (1995). A procedural explanation of the generation effect: The use of an operand retrieval strategy for multiplication and addition problems. *Journal of Memory & Language*, **34**, 399-416.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior*, **16**, 519-533.

Museum of Science, Boston (2004). *Leonardo da Vinci, Renaissance man*. Retrieved September 15, 2004 from www.mos.org/leonardo/bio.html.

Nairne, J. S., & Widner, R. L. (1987). Generation effects with nonwords: The role of test appropriateness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 164-171.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, **95**, 109-133.

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.

Rudell, R. B., & Speaker, R. B., Jr. (1985). The interactive reading process: A model. In H. Singer & R. B. Rudell (Eds.), *Theoretical models and processes of reading* (3rd ed., pp. 751-793). Newark, DE: International Reading Association.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 204-221.

State of Hawaii, Division of Aquatic Resources, Department of Land and Natural Resources (2004). *Sharks of Hawaii.* Retrieved September 15, 2004, from www.aloha.com/~lifeguards/sharintr.html#forty.

Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, **28**, 129-160.

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

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.

Weaver, C. A., III (1990). Constraining factors in calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, & 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.

Weaver, C. A., III, & Kelemen, W. L. (2003). Processing similarity does not improve metamemory: Evidence against transfer-appropriate monitoring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 1058-1065.

Zacks, R. T. (1969). Invariance of total learning time under different conditions of practice. *Journal of Experimental Psychology*, **82**, 441-447.

NOTE

1. The authors acknowledge that the differences in cued-recall performance magnitude between Experiments 1 and 2 could be influenced by scale dependence.

(Manuscript received September 8, 2005;
revision accepted for publication April 9, 2006.)