Metacognitive Control: Children's Short-Term Versus Long-Term Study Strategies

LISA K. SON
Department of Psychology
Barnard College

ABSTRACT. First graders were tested on 2 tasks investigating metacognitive control abilities. In Experiment 1, after studying and making judgments of learning (JOLs) on word–synonym pairs, they could either mass or space each pair's restudy trials. If they chose to mass study, then the pair was presented again immediately. If they chose to space study, then the pair was presented again after a delay, testing a long-term control strategy. Results showed that children's JOLs did not influence study strategy. Furthermore, in general, children chose to mass more often than to space. In Experiment 2, after studying and making JOLs on word–synonym pairs, they could either read the entire pair again or they could generate the target by attempting to retrieve the target synonym when shown only the word. Unlike Experiment 1, both read and generate decisions were carried out in the short term. Also unlike Experiment 1, Experiment 2's results showed that the higher the child's JOL, the more likely they chose generation strategies—showing evidence for short-term metacognitively controlled strategies.

Key words: children's metacognitive control, judgments of learning, metacognitive control, short-term and long-term control

IT IS EASY TO UNDERSTAND why the optimization of long-term learning is a critical issue in the field of cognitive psychology. As educators, teachers strive for their students not merely to learn and store information for the short term, but to constantly remember and build on that knowledge far into the future. As learners in school, students also study with the aim to recall the studied information,

This research was supported by start-up funds from the Department of Psychology of Barnard College. The author thanks Noa Besner, Mary Bruch, Molly Flaherty, Emily Freeman-Daniels, Diana Lee, Danielle Sussan, Stephanie Tsai, and Jacqueline Yunits for help in collection of the data.

Address correspondence to Lisa Son, Department of Psychology, Barnard College, 3009 Broadway, New York, NY 10027; lson@barnard.edu (e-mail).

if not years into the future, at least during a test given the following day, week, or month. To be a successful learner, then, the ability to control one's study so as to optimize long-term performance is crucial. Many data, however, have suggested that making good study decisions that will benefit future long-term performance may be difficult. In this article, several reasons for why this is so are presented. In particular, the inconsistencies between short-term memory and long-term memory will be discussed and supported with data testing metacognitive control in first-grade children.

Researchers have long thought that the ability to make good study decisions depends on a person's ability to accurately assess the learnability of the to-be-learned materials (A. L. Brown, 1987; Flavell, 1976, l979; Flavell & Wellman, l977; Hart, 1965; Underwood, 1966). Nelson and Narens (1990, 1994) formally proposed this concept in a framework describing a monitoring process and a control process. Their monitoring process consists of a meta-level component that judges how well learned an item is at the cognitive level. For example, when studying A, B, and C, the learner might judge A as very difficult, B as somewhat easy, and C as completely learned. There has been an outpouring of scientific work on the accuracy of these metacognitive judgments, with generally positive results—many have found that people are relatively good at monitoring their learning (Dunlosky & Nelson, 1992, 1994, 1997; King, Zechmeister, & Shaughnessy, 1980; Koriat, 1975, 1993, 1995, 1997, 1998; Koriat & Goldsmith, 1996, 1998; Leonesio & Nelson, 1990; Lovelace, 1984; Mazzoni, Cornoldi, Tomat, & Vecchi, 1997; Mazzoni & Nelson, 1995; Metcalfe, Schwartz, & Joaquim, l993; Metcalfe & Weibe, l987; Nelson, 1988; Nelson & Dunlosky, 1991, 1994; Nelson, Leonesio, Landwehr, & Narens, 1986; Schwartz & Metcalfe, l994; Schwartz & Smith, 1997; Smith, Brown, & Balfour, 1991; Thiede & Dunlosky, 1994; Vesonder & Voss, 1985).

Nelson and Narens' control process (1990, 1994) can be described as one in which people can use their judgments to make various study decisions. Whether the decisions are optimal, at present, is complicated. For example, having judged A as difficult, B as somewhat easy, and C as completely learned, should the most study time be allocated to A? And over the next hour, how many times and in what schedules should A be reviewed? And, should C be ignored from now until test? Studies investigating such control questions have been scrutinized a great deal recently, and new paradigms have been designed to test decisions such as those of study-time allocation (Dunlosky & Hertzog, 1998; Mazzoni & Cornoldi, 1993; Mazzoni, Cornoldi, & Marchitelli, 1990; Metcalfe, 2002; Metcalfe & Kornell, 2003; Nelson, Dunlosky, Graf, & Narens, 1994; Nelson & Leonesio, 1988; Nelson & Narens, 1990, 1994; Son & Metcalfe, 2000; Thiede & Dunlosky, 1994), item selection (Thiede & Dunlosky, 1999), and scheduling of study—that is, massing or spacing (Son, 2004). The current research focuses on *scheduling* of study as well as self-testing or *generation* during study—scheduling being a test of long-term control and generation being a test of short-term control. The data will suggest that although young children use metacognitive strategies when the choices they make pan out in the short term, they do not use those strategies when the choices do not pan out until the long term. First, however, a case is made for why long-term processes are difficult to take control of, particularly in light of people's attention to the short term.

Short-term memory and long-term memory have classically been divided into separate stores, subsystems, or systems (Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974; Broadbent, 1958; J. Brown, 1958; Craik, 1970; Gardiner, Craik, & Bleasdale, 1973; Miller, 1956; Murdock, 1974; Peterson & Peterson, 1959; Schacter & Tulving, 1994; Shiffrin, 1973; Waugh & Norman, 1965), or as having separate retrieval processes (Tulving, 1967). One example of their incompatible nature is when items that were once remembered the best in short-term memory undergo an unexpected and opposite "fate" (Craik, p. 143) in the long term. In Craik's experiment, although people remembered the most recent items from a list the best during an immediate test, they remembered those exact items the least best during a test given in the long term. Soon thereafter, Gardiner et al. demonstrated a similar finding in which words retrieved with difficulty during a general-knowledge question–answering task were later remembered more easily than were words that had been retrieved more easily on the initial question-answering task. These findings are similar to those of the generation effect, which reveals that generating, or testing oneself, during study (in the short term), although likely to be more difficult and error-provoking than passively reading, leads to higher long-term test performance than that following passive reading (Gardiner & Hampton, 1985; Gardiner & Rowley, 1984; Graf, 1980; Jacoby, 1978; Johns & Swanson, 1988; Nairne & Widner, 1987; Slamecka & Graf, 1978). The spacing effect, in which spacing leads to significantly higher long-term performance than does massing (Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Bahrick & Phelps, 1987; Bjork & Allen, 1970; Dempster, 1987, 1988; Glenberg, 1976, 1977, 1979; Glover & Corkill, 1987; Hintzman, 1969, 1974; Landauer, 1967; Melton, 1970; Shaughnessy, Zimmerman, & Underwood, 1972; Toppino, 1991, 1993; Toppino & DiGeorge, 1984; Underwood, 1970; Zechmeister & Shaughnessy, 1980), also is assumed to occur because of the effortful, active, retrieval process people engage in during spacing, but not massing (Birnbaum & Eichner, 1971; Bjork, 1975; Donovan & Radosevich, 1999; Glenberg, 1976, 1977, 1979; Glenberg & Smith, 1981; Glover, 1989; Melton, 1970). These studies, as well as others, paint an incongruous picture of the nature of one's memory for an item in the short term and memory for that same item in the long term.

The above incompatibilities between short-term performance and long-term performance suggest that people may have trouble monitoring their learning. For example, given the above incongruities, if A is remembered best now, it may be at risk of being judged as well learned, but then being forgotten at a long-term test. This hypothesis is consistent with the notion that people access information simultaneously from both long-term and short-term memory, potentially resulting in short-term memory interference of long-term performance assessment (Mohs, Wescourt, & Atkinson, 1973). In fact, researchers have found that people make inaccurate metacognitive judgments, particularly when asked to make judgments about long-term performance (Begg, Snider, Foley, & Goddard, 1989; Benjamin, Bjork, & Schwartz, 1998; Kelemen & Weaver, 1997; Nelson & Dunlosky, l991; Weaver & Kelemen, 1997; Zechmeister & Shaughnessy, 1980). For example, Zechmeister and Shaughnessy showed that people reported higher confidence judgments during massed study than during spaced study. Of course, at test, memory for the massed items was worse than memory for the spaced items. Benjamin et al. showed that people's judgments increased the more fluently an item was retrieved in the short term. However, those judgments were in contrast to what was actually remembered in the long term. A similar finding was obtained by Begg et al., who showed that participants gave higher predictions of later recognizability to high-frequency words than to low-frequency words, despite the superiority of low-frequency word recognition.

The most apparent data supporting the incongruity between short-term and long-term monitoring have been those of the delayed–judgments-of-learning (JOL) effect. JOLs are metacognitive judgments that are made predicting future long-term test performance. Researchers have shown that when people are asked to make their JOLs immediately after study, the JOLs are not nearly as accurate as they are when asked to make their JOLs after a delay. This effect has been found in both adults (Kelemen & Weaver, 1997; Kimball & Metcalfe, 2003; Nelson & Dunlosky, l991; Weaver & Kelemen, 1997) and in children (Koriat & Shitzer-Reichert, 2002). Several explanations have been proposed for the delayed-JOL effect. Nelson and Dunlosky proposed what is called the Monitoring-Dual-Memories Hypothesis, which assumes that the JOLs are made by retrieving information, about their learning of a particular item, from both short-term memory and long-term memory. When making immediate JOLs, information from the short term is accessible and dominant. By contrast, when making delayed JOLs, the information will be based primarily on long-term memory, with little input from short-term memory. Kelemen and Weaver went further, suggesting that eliminating progressively larger amounts of short-term memory contamination should progressively increase JOL accuracy. Another explanation (Begg et al., 1989; Glenberg, 1987) proposes that the delayed-JOL effect occurs because of differences in the degree of contextual match from the time of the judgment (during short-term memory) to the time of the test (during long-term memory), especially in the case of an immediate JOL. In short, the delayed-JOL effect suggests that during study, if people monitor their learning immediately, then their judgments become less accurate in predicting long-term performance because of the bombardment of information from the short term and the inadvertent but inevitable unawareness of long-term changes.

Going hand in hand with the problem of inaccurate monitoring, the question that is posed here is whether people use their judgments to control study in ways that would benefit only short-term performance, or whether they use their judgments in ways that might benefit long-term performance. Recently, it had been found that, in fact, adults systematically use their JOLs to control their spacing strategies, and, more interestingly, they do not always choose to optimize merely short-term performance (Son, 2004). In fact, people choose to space most of their study sessions rather than to mass. More specifically, people choose to mass items that are judged as difficult and space items that are judged as easy, which supports the assumption that people choose to study an item until it is fully encoded, and only until fully encoded (Bahrick et al., 1993; Simon & Bjork, 2001). As a comparison, in Experiment 1 of the present study, first-grade children were tested on a similar control-of-spacing task, after which, having made JOLs, they were given a choice to mass their study (favoring short-term performance) or to space their study (favoring long-term performance). In Experiment 2 of the present study, children were tested on a control-of-generation task in which the choice was between reading the item again (favoring short-term performance) or generating the item by testing oneself on that item (favoring long-term performance).

EXPERIMENT 1

In Experiment 1, children's spacing strategies were tested. There have been a few published studies investigating the spacing effect in children (Cahill & Toppino, 1993; Rea & Modigliani, 1987; Toppino, 1991, 1993; Toppino & DiGeorge, 1984). None, however, has tested children's choices of spacing. After making JOLs on various vocabulary word–synonym pairs, they were given the option of massing or spacing their study. On the one hand, selecting the option to mass would require that the child make a study decision that would be resolved in the short term. On the other hand, selecting the option to space would require that the child make a study decision that would be resolved in the long term—the second study session would occur after a delay. The critical questions were to see whether children would control their study in such a way as to (a) space their study—seemingly a more optimal strategy than is massing study, according to the spacing effect, and (b) incorporate their judgments when making massing or spacing decisions.

Method

Participants

All participants were first-grade students from an elementary school in Manhattan, P.S.75. To become a participant, each child's parent read and signed a consent form explaining all methods, procedures, benefits, and risks of the study, adhering to American Psychological Association guidelines. The final pool of participants consisted of 11 boys and 13 girls. Their ethnic backgrounds were as follows: 10 Hispanic, 5 African American, 5 Caucasian American, 1 Asian American, and 3 American Indian children. All were either 6 or 7 years of age.

Materials

The stimuli were 30 word–synonym pairs, randomly selected from a pool of 100 pairs, taken from vocabulary workbooks used in grades 1–3 so that there were items that would be both easy and hard for the children to learn. An example of an easy item was "angry–mad." An example of a difficult item was "occupation–job." For each child, the computer program randomly selected the 30 items to study.

Procedure

Each child was placed in front of a computer and told that they would be reading some words and studying their meanings. They were also told that they would be tested on the words later on, in about 10 min. An adult was seated with them throughout the length of the experiment. Before beginning, each child was asked if they knew what a synonym was and to give an example of a synonym for "large." After it was clear that they understood the synonym concept, a practice session began. During the practice session, eight word–synonym pairs were presented on the screen, one at a time, for 3 s each. Each child was asked to read the words while trying to remember the meaning. After each presentation, they made a JOL of how well they thought that they would be able to remember the meaning of the word, that is, "mad" if they were given "angry" later on during the memory test. The judgments were made by manipulating a pointer, via the mouse, to a particular place on a slider presented on the screen. They were told that moving the pointer to the very far right meant that they were very sure that they would remember the meaning, moving the pointer to the very far left meant that they were sure that they would forget the meaning, and that they could also move the pointer to somewhere in the middle if they were not completely sure or unsure. The slider values ranged from 0 to 100, although the actual numbers were not displayed. The main goal of the practice session was to make sure that the children understood the task so that they could practice using the slider to make their JOLs.

Immediately after the practice session, each child was presented with a list of 30 word–synonym pairs to study for a later test. As in the practice session, each pair was presented for 3 s. During those 3 s, the child was asked to read the pair out loud. After each presentation, the child would make JOLs on the slider. After making the judgment for each item, the children were then given two options on the screen: (a) They could choose to see the same pair in a massed session—indicated by a *Study Now* button, or (b) they could choose to study that pair again in a spaced session—indicated by a *Study Later* button.[1] If *Study Now* (Massed) was chosen, then the pair was presented again immediately for 3 s. If *Study Later* (Spaced) was chosen, then that pair was shown again (for 3 s) after the entire list had been presented. After the entire list had been presented—some massed and some spaced—there was a 1-min distracter task in which children were asked to count to 30 twice. Finally, there was a cued-recall test in which only the word from each pair was presented. The child's task was to say the synonym for each word, to be typed in by the experimenter.

Results

The JOL data were analyzed using normalized scores, divided into 3 levels (lowest one third, middle one third, and highest one third) for each child. The reason for the three categories was that most of the children seemed to choose either a low, middle, or high rating for their judgments, rather than using the entire range of the slider. This may have also been a result of the nature of the instructions, in which very high judgments, very low judgments, and judgments somewhere in the middle had been emphasized. For the statistical analyses, a probability level of $p < .05$ was used as the criterion for statistical significance and a level of $p < .10$ as the criterion for a marginal trend. Estimates of effect size were calculated with Cohen's *d*.

To get a sense of whether the children's metacognitive judgments were accurate, gamma correlations (Nelson, 1984) between JOL and final test performance were first calculated. Because the JOLs were made immediately after study, rather than after a delay, and various study sessions occurred before the final test, the correlations were not expected to be enormous. Still, the mean gamma was significantly positive ($M = 0.17$), $t(23) = 2.75$, $SE = 0.06$, suggesting that the children were, at least, making judgments that were fairly logical.

Son (2004) found that adults' spacing strategies depended on their judgments, such that the higher an item's judgment, the more likely it would be that the item would be chosen for a spaced study repetition. To see if the children behaved in the same way, proportions of items that were chosen as massed and spaced for each of the three judgment levels were computed. An analysis of variance (ANOVA) with spacing strategy (massed, spaced) and JOL level (low, middle, high) resulted in a marginal effect of spacing strategy, $F(1, 15) = 2.95$, $MSE = 2.07$, $d = 0.16$, $p = .10$, which indicated that children seemed to lean toward the massing strategy rather than the spacing strategy, particularly at the medium judgment level, $t(15) = 2.24$, $SE = 0.21$. There was no interaction with JOL level, indicating that the judgments did not affect the children's spacing decisions. The results are shown in Figure 1.

Discussion

The results from Experiment 1 show that children did not use their metacognitive judgments to select their spacing strategies. And, significantly at the middle judgment level, they chose to mass more often than they chose to

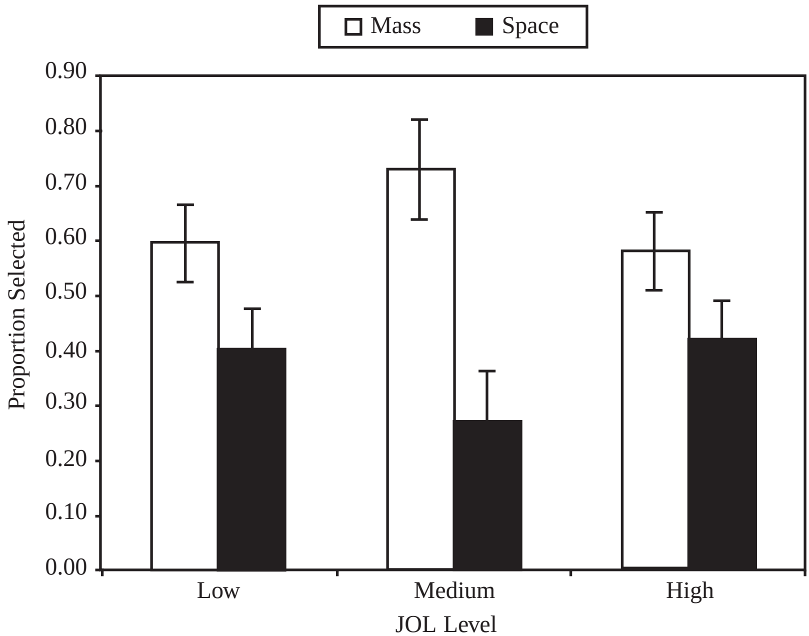

FIGURE 1. Bars represent the mean proportion of massed (white bars) and spaced (black bars) items across JOL level (*z* scores calculated for each participant and split into 3 levels from least confident to most confident).

space. One could interpret the data by saying that children are not able to make study decisions that benefit long-term learning, such as spacing. Or that they do not yet possess the ability to control study based on their metacognitions. In the developmental literature, however, others have found evidence that children were able to make systematic decisions based on the learnability of an item (Dufresne & Kobasigawa, 1988, 1989; Kobasigawa & Metcalf-Haggert, 1993). For example, Kobasigawa and Metcalf-Haggert found that even first graders choose to allocate more study time to materials that are more difficult. The difference between their results and the results of Experiment 1 might be attributed to the fact that spaced sessions occur in the long term, as well as the feedback that accompanies those decisions. In a study-time allocation task, however, the decisions are executed during short-term study, whether one chooses to study a lot or a little. A question that remains, then, is whether children could make metacognitively controlled decisions that still benefit long-term performance, so long as those decisions pan out in the short term. This question was explored in Experiment 2.

EXPERIMENT 2

Given that the children did not use metacognitively controlled strategies in Experiment 1, a simpler task was conducted in Experiment 2. The *generation effect* states that generating an item leads to better long-term performance than passively reading an item (Gardiner & Hampton, 1985; Gardiner & Rowley, 1984; Graf, 1980; Jacoby, 1978; Johns & Swanson, 1988; Nairne & Widner, 1987; Slamecka & Graf, 1978). Generation is encouraged when students are given frequent tests and pop quizzes, for example, when they are encouraged to generate or present their own ideas. Data have also shown that the mere act of retrieval induced by a recall test can be considered more potent than a passive study opportunity in facilitating long-term recall (Bjork, 1988; Landauer & Bjork, 1978). However, although a few studies have tested the generation effect in children (Calvert, 1991; McFarland, Duncan, & Bruno, 1983), there have been no data testing whether young learners use any systematic learning strategies in regard to the generation of information. In Experiment 2, after making JOLs on vocabulary pairs, children were given the choice of either reading the item again (which would be suboptimal for long-term performance according to the generation effect) or of generating the item themselves (which would be beneficial for long-term performance). At the same time, like study-time allocation procedures, both reading and generation decisions, along with feedback, were carried out in the short term.

Method and Procedure

The participants were a group of 27 first graders, 15 girls and 12 boys, some of whom were tested in Experiment 1, and some who were not. Their ethnic backgrounds were as follows: 10 Hispanic, 11 African American, 4 Caucasian American, 1 Asian American, and 1 American Indian children. The procedure was similar to that of Experiment 1. Children were told that they would be studying various pairs that would be tested later on in about 10 min. After a practice session, the participants were presented with a list of 35 word–synonym pairs (taken from the same 100 word-pair pool), such as “angry–mad.” After each presentation, they made a JOL using the slider. Then, they were asked whether they wanted to read or whether they wanted to generate the synonym when given the word. If they chose to read the word, then the entire word–synonym pair appeared on the screen for 3 s. If they chose to generate the synonym, then only the word appeared on the screen again, and they were asked to say the synonym that was related to the word to be typed in by the experimenter. After each generated item, correct feedback was presented for 1 s. Thus, both the read and generate events, and their feedback, occurred in the short term. After studying the entire list, including read and generated items, there was a math distracter task for 1 min, followed by a cued-recall test.

Results

As in Experiment 1, each child's JOLs were divided into three categories: low, middle, and high, based on normalized judgments. Proportions of choices for read and generate were then calculated. The results are presented in Figure 2. As can be seen in the figure, the choices depended significantly on their JOL level. The interaction between JOL level and study strategy was significant, $F(2, 10) = 9.88$, $MSE = 1.18$, $d = 0.66$. For low JOLs, they chose to read the items more often, $t(26) = 9.26$, $SE = 0.07$. For high JOLs, they chose to generate the items on their own, $t(26) = 6.20$, $SE = 0.10$. The proportions of read and generate choices were not different at the middle JOL level. These data are in contrast to those of Experiment 1, and provide some optimism for the notion that children can make study decisions that are metacognitively guided, and optimal for long-term performance.

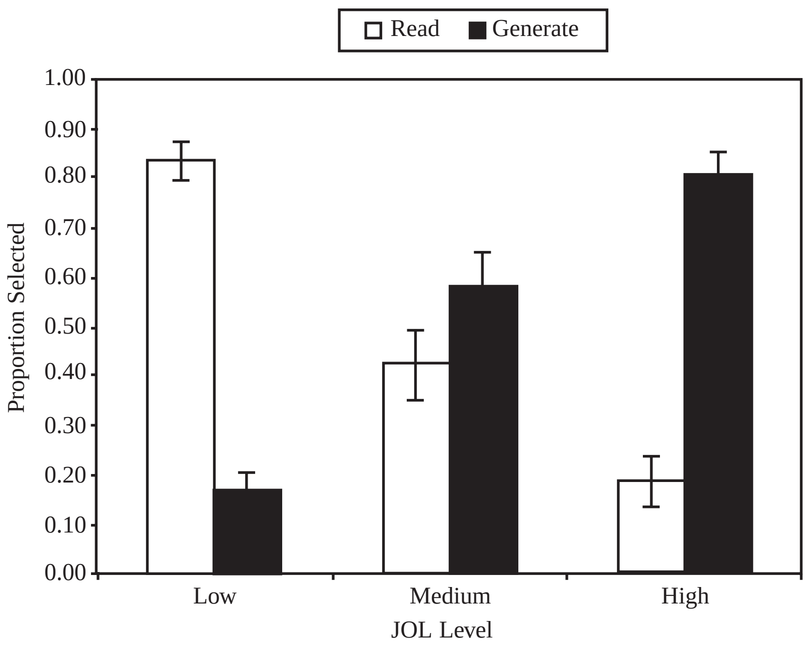

FIGURE 2. Bars represent the mean proportion of read (white bars) and generated (black bars) items across JOL level (*z* scores calculated for each participant and split into 3 levels from least confident to most confident).

GENERAL DISCUSSION

In Experiment 1, children chose massing rather than spacing strategies, and they chose those strategies while ignoring their own metacognitive judgments. However, previous research had found that children, even in the first grade, are able to make effective study decisions, such as those of study-time allocation (Dufresne & Kobasigawa, 1989; Kobasigawa & Metcalf-Haggert, 1993). One possible discrepancy between those experiments and Experiment 1 is that in previous studies, all study-time–allocation decisions were implemented in the short term, resulting in immediate feedback. For example, if a particular item was difficult, choosing to spend a short time or a long time re-studying that item would be two different decisions with two different outcomes, but would still both occur at the present time, during short-term processing. In Experiment 1, however, only the decision to mass one's study would actually pan out in the short term. Decisions to space one's study would not be executed until a delay had passed, in the long term.

In Experiment 2, a similar study was conducted, except that, as in previous studies, the study options that the children chose were implemented with feedback during the short term. Here, children had to select between reading the item passively or generating the item actively, which would seemingly be more beneficial for long-term performance. Now, children were able to make some optimal decisions. They did not merely passively read the to-be-learned items, but sometimes chose to generate the items on their own—a more difficult and active strategy in the short term than reading. In addition, the children's strategies were guided by their metacognitive judgments: They chose to generate items that were given high JOLs and chose to read items that were given low JOLs. These data provide evidence of good control abilities in young children, and of abilities to make decisions that would be beneficial for long-term performance.

As described in the introduction of this article, during study, people are highly at risk of experiencing illusions of learning owing to the bombardment of information from short-term process. As a result of this heightened awareness of processes occurring in the short term, people's study decisions are prone to rely on judgments made in the short term, as well as on any kind of feedback that might be consumed in the short term, especially if previous experience from long-term knowledge sources is lacking. In Experiment 1, it is likely that the children were relying on feedback during study, which led to a greater amount of massing strategies than spacing strategies, where feedback would not occur until long after the choice had been made. And essentially, the ability to know that ceasing study now only to re-study the item during a spaced session in the long-term future would only come after having experienced the benefits of spacing from previous encounters, which first graders have had little of. Indeed, Tulving (1983, 1984) was the one who stated that people do have the ability to mentally time travel, but that directing oneself into an as-yet nonexisting future could not occur without an ability to reflect on a past. After children have a past to reflect on—as when older—or they are given constant short-term feedback—as was the case in Experiment 2—then suddenly, their study decisions are guided skillfully by their metacognitive judgments.

Simply possessing awareness of an item's learning state would be of no value without using that knowledge to control subsequent study behavior (A. L. Brown, 1987; Flavell, 1976; Kluwe, 1982; Nelson & Narens, 1990, 1994). And using that knowledge inappropriately or nonoptimally would have just as little value. Learners need to avoid misperceptions of passive and fluent strategies, which benefit only short-term performance, and instead shift to active and effortful strategies, which are optimal for long-term performance. In this article, the two main questions were (a) to see if children made systematic choices that would be appropriate for benefiting long-term performance, and (b) to see if those study choices were guided by their own metacognitive judgments. In the end, the results showed that children can make some metacognitive decisions that benefit long-term performance, but that those abilities are not fully developed for decisions that necessitate awareness of long-term future outcomes, such as the decision to space study. How awareness of long-term control strategies in children might be advanced and optimized is left for future research.

NOTE

1. When adults were tested on the same task (Son, 2004), they had also been given a *done* button, allowing them to decide not to re-study that particular item. I had used the *done* button when testing the children originally. However, the children chose that option on a majority of the trials, indicating that they were not motivated to select massing or spacing strategies. Thus, I removed that option.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–195). New York: Academic Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8). New York: Academic Press.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science, 4,* 316–321.

Bahrick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 344–349.

Begg, I., Snider, A., Foley, F., & Goddard, R. (1989). The generation effect is no artifact: Generating makes words distinctive. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 977–989.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127,* 55–68.

Birnbaum, I. M., & Eichner, J. T. (1971). Study versus test trials and long-term retention in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 10,* 516–521.

Bjork, R. A. (l975). Short-term storage: The ordered output of a central processor. In F.

Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory* (pp. 151–171). Hillsdale, NJ: Erlbaum.

Bjork, R. A. (l988). Retrieval practice. In M. Gruneberg & P. E. Morris (Eds.), *Practical aspects of memory: Current research and issues, Vol. 1: Memory in everyday life* (pp. 396–401). New York: Wiley.

Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding. *Journal of Verbal Learning and Verbal Behavior, 9,* 567–572.

Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon.

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65–116). Hillsdale, NJ: Erlbaum.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Journal of Experimental Psychology, 10,* 2–21.

Cahill, A., & Toppino, T. C. (1993). Young children's recognition as a function of the spacing of repetitions and the type of study and test stimuli. *Bulletin of the Psychonomic Society, 31,* 481–484.

Calvert, S. L. (1991). Presentational features for young children's production and recall of information. *Journal of Applied Developmental Psychology, 12,* 367–378.

Craik, F. I. M. (1970). The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior, 9,* 143–148.

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79,* 162–170.

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist, 43,* 627–634.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84,* 795–805.

Dufresne, A., & Kobasigawa, A. (1988). Developmental differences in children's spontaneous allocation of study time. *Journal of Genetic Psychology, 149,* 87–92.

Dufresne, A., & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: Differential and sufficient aspects. *Journal of Experimental Child Psychology, 47,* 274–296.

Dunlosky, J., & Hertzog, C. (l998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249–276). Mahwah, NJ: Erlbaum.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20,* 374–380.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language, 33,* 545–565.

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 and Language, 36,* 34–49.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231–235). Hillsdale, NJ: Erlbaum.

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

Flavell, J. H., & Wellman, H. M. (l977). Metamemory. In R. V. Kail & J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 3–33). Hillsdale, NJ: Erlbaum.

Gardiner, J. M., Craik, F. I., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition, 1,* 213–216.

Gardiner, J. M., & Hampton, J. A. (1985). Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 732–741.

Gardiner, J. M., & Rowley, J. M. (1984). A generation effect with numbers rather than words. *Memory & Cognition, 12,* 443–445.

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior, 15,* 1–16.

Glenberg, A. M. (1977). Influences of retrieval processes on the spacing effect in free recall. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 282–294.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7,* 95–112.

Glenberg, A. M. (1987). Temporal context and recency. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus Centennial Conference* (pp. 173–190). Hillsdale, NJ: Erlbaum.

Glenberg, A. M., & Smith, S. M. (1981). Spacing repetitions and solving problems are not the same. *Journal of Verbal Learning and Verbal Behavior, 20,* 110–119.

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81,* 392–399.

Glover, J. A., & Corkill, A. J. (1987). Influence of paraphrased repetitions on the spacing effect. *Journal of Educational Psychology, 79,* 198–199.

Graf, P. (1980). Two consequences of generating: Increased inter- and intraword organization of sentences. *Journal of Verbal Learning and Verbal Behavior, 19,* 316–327.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56,* 208–216.

Hintzman, D. L. (1969). Recognition time: Effects of recency, frequency and the spacing of repetitions. *Journal of Experimental Psychology, 79,* 192–194.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77–99). Hillsdale, NJ: Erlbaum.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17,* 649–667.

Johns, E. E., & Swanson, L. G. (1988). The generation effect with nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 180–190.

Kelemen, W. L., & Weaver, C. A. III, (1997). Enhanced metamemory at delays: Why do judgments of learning improve over time? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1394–1409.

Kimball, D. R., & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory & Cognition, 31,* 918–929.

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.

Kluwe, R. H. (1982). Cognitive knowledge and executive control. In D. Griffin (Ed.), *Human mind–Animal mind* (pp. 201–224). New York: Springer.

Kobasigawa, A., & Metcalf-Haggert, A. (1993). Spontaneous allocation of study time by first- and third-grade children in a simple memory task. *Journal of Genetic Psychology, 154,* 223–235.

Koriat, A. (1975). Phonetic symbolism and feeling of knowing. *Memory & Cognition, 3,* 545–548.

Koriat, A. (1993). How do we know what we know? The accessibility model of the feeling of knowing. *Psychological Review, 100,* 609–639.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124,* 311–333.
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.
Koriat, A. (1998). Metamemory: The feeling of knowing and its vagaries. In M. Sabourin & F. Craik (Eds.), *Advances in psychological science: Vol. 2. Biological and cognitive aspects* (pp. 461–479). Hove, England: Psychology Press.
Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490–517.
Koriat, A., & Goldsmith, M. (1998). The role of metacognitive processes in the regulation of memory performance. In G. Mazzoni & T. O. Nelson (Eds.), *Metacognition and cognitive neuropsychology: Monitoring and control processes* (pp. 97–118). Mahwah, NJ: Erlbaum.
Koriat, A., & Shitzer-Reichert, R. (2002). Metacognitive judgments and their accuracy. In P. Chambres & M. Izaute (Eds.), *Metacognition: Process, function and use* (pp. 1–17). Dordrecht, The Netherlands: Kluwer Academic Publishers.
Landauer, T. K. (1967). Interval between item repetitions and free recall memory. *Psychonomic Science, 8,* 439–440.
Landauer, T. K., & Bjork, R. A. (1978). Optimal rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.
Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 464–470.
Lovelace, E. A. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 756–766.
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.
Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1263–1274.
McFarland, C. E., Duncan, E., & Bruno, J. M. (1983). Developmental aspects of the generation effect. *Journal of Experimental Child Psychology, 36,* 413–428.
Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 596–606.
Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General, 131,* 349–363.
Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General, 132,* 530–542.
Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 851–864.
Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition: Knowing about knowing*. Cambridge, MA: MIT Press.
Metcalfe, J., & Weibe, D. (1987). Intuition in insight and noninsight problem solving.

Memory & Cognition, 15, 238–246.
Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.
Mohs, R. C., Wescourt, K. T., & Atkinson, R. C. (1973). Effects of short-term memory contents on short- and long-term memory searches. *Memory & Cognition, 1,* 443–448.
Murdock, B. B., Jr. (1974). *Human memory: Theory and data*. Hillsdale, NJ: Erlbaum.
Nairne, J. S., & Widner, R. L. (1987). Generation effects with nonwords: The role of test appropriateness. *Journal of Experimental Psychology: Learning, Memory, and 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. (1988). Predictive accuracy of the feeling of knowing across different criterion tasks and across different subject populations and individuals. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (pp. 190–196). Chichester, England: Wiley.
Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOL) are extremely accurate at predicting subsequent recall: The delayed-JOL effect. *Psychological Science, 2,* 267–270.
Nelson, T. O., & Dunlosky, J. (1994). How shall we explain the delayed-judgment-of-learning effect? *Psychological Science, 3,* 317–318.
Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5,* 207–213.
Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 676–686.
Nelson, T. O., Leonesio, R. J., Landwehr, R. S., & Narens, L. (1986). A comparison of three predictors of an individual's memory performance: The individual's feeling of knowing versus the normative feeling of knowing versus base-rate item difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 279–287.
Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–141) New York: Academic Press.
Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–25). Cambridge, MA: MIT Press.
Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual items. *Journal of Experimental Psychology, 61,* 12–21.
Rea, C. P., & Modigliani, V. (1987). The spacing effect in 4- to 9-year-old children. *Memory & Cognition, 15,* 436–443.
Schacter, D. L., & Tulving, E. (1994). What are the memory systems of 1994? In D. L. Schacter & E. Tulving (Eds.), *Memory systems* (pp. 1–38). Cambridge, MA: MIT Press.
Schwartz, B. L., & Metcalfe, J. (1994). Methodological problems and pitfalls in the study of human metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 93–113). Cambridge, MA: MIT Press.
Schwartz, B. L., & Smith, S. M. (1997). The retrieval of related information influences tip-of-the-tongue states. *Journal of Memory & Language, 36,* 68–86.
Shaughnessy, J. J., Zimmerman, J., & Underwood, B. J. (1972). Further evidence on the MP–DP effect in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 11,* 1–12.
Shiffrin, R. M. (1973). Information persistence in short-term memory. *Journal of Experi-*

mental Psychology, 100, 39–49.
Simon, D., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 907–912.
Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory, 4,* 592–604.
Smith, S. M., Brown, J. M., & Balfour, S. P. (1991). TOTimals: A controlled experimental method for studying tip-of-the-tongue states. *Bulletin of the Psychonomic Society, 29,* 445–447.
Son, L. K. (2004). Metacognitively controlled spacing of study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 601–604.
Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 204–221.
Thiede, K. W., & Dunlosky, J. (1994). Delaying students' metacognitive monitoring improves their accuracy in predicting their recognition performance. *Journal of Educational Psychology, 86,* 290–302.
Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1024–1037.
Toppino, T. C. (1991). The spacing effect in young children's free recall: Support for automatic-process explanations. *Memory & Cognition, 19,* 159–167.
Toppino, T. C. (1993). The spacing effect in preschool children's free recall of pictures and words. *Bulletin of the Psychonomic Society, 31,* 27–30.
Toppino, T. C., & DiGeorge, W. (1984). The spacing effect in free recall emerges with development. *Memory & Cognition, 12,* 118–122.
Tulving, E. (1967). The effects of presentation and recall of material in free recall learning. *Journal of Verbal Learning and Verbal Behavior, 6,* 175–184.
Tulving, E. (1983). *Elements of episodic memory*. Oxford, NY: Oxford University Press.
Tulving, E. (1984). Multiple learning and memory systems. In K. M. J. Lagerspetz & P. Niemi (Eds.), *Psychology in the 1990s* (pp. 163–184). Amsterdam: Elsevier.
Underwood, B. J. (1966). Individual and group predictions of item difficulty for free-recall learning. *Journal of Experimental Psychology, 71,* 673–679.
Underwood, B. J. (1970). A breakdown of the total-time law in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 9,* 573–580.
Vesonder, G., & Voss, J. (1985). On the ability to predict one's own responses while learning. *Journal of Memory and Language, 24,* 363–376.
Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review, 72,* 89–104.
Weaver, C. A., III, & Kelemen, W. L. (1997). Judgments of learning at delays: Shifts in response patterns or increased metamemory accuracy? *Psychological Science, 8,* 318–321.
Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society, 15,* 41–44.

Manuscript accepted January 2005