

Study Efficacy and the Region of Proximal Learning Framework

Nate Kornell and Janet Metcalfe
Columbia University

One of the most important reasons to investigate human metacognition is its role in directing how people study. However, limited evidence exists that metacognitively guided study benefits learning. Three experiments are presented that provide evidence for this link. In Experiment 1, participants' learning was enhanced when they were allowed to control what they studied. Experiments 2a–d replicated this finding and showed contributions of self-regulated study to learning. Experiments 3a and 3b showed that, when forced to choose among items they did not know, participants chose the easiest items and benefited from doing so, providing evidence for the link between metacognitive monitoring/control and learning, and supporting the region of proximal learning model of study-time allocation.

Keywords: metacognition, study-time allocation, region of proximal learning, education, memory

One of the best reasons to study metacognition is because it has the potential to play a large part in guiding how people study and, as a result, in how effectively they learn (e.g., Benjamin, Bjork, & Schwartz, 1998; Dunlosky & Hertzog, 1997; Metcalfe, 2002; Metcalfe & Kornell, 2003, 2005; T. O. Nelson & Narens, 1990, 1994). People use memory monitoring, especially judgments of learning (JOLs), to decide which items to study and how long to spend on them (e.g., Mazzoni, Cornoldi, & Marchitelli, 1990; Metcalfe, 2002; T. O. Nelson, Dunlosky, Graf, & Narens, 1994; T. O. Nelson & Leonesio, 1988; T. O. Nelson & Narens, 1990; Son & Metcalfe, 2000). The central question addressed here is, does such metacognitively guided study lead to effective learning? The answer can be arrived at by asking more basic questions. What *should* people choose to study? What *do* people choose to study? And, are they the same? The answers to these questions have implications for pedagogy, how learning works, and the debate over the accuracy (or lack thereof) of metacognitive monitoring and control as well as theoretical implications for current models of study-time allocation.

Until recently, the dominant model of study-time allocation has been discrepancy reduction (DR) (e.g., Dunlosky & Hertzog, 1998). According to this model, a person chooses to study the items that are farthest from being learned, that is, the most difficult items for that person (e.g., Dunlosky & Hertzog, 1998; Dunlosky & Thiede, 1998; T. O. Nelson & Narens, 1990; Thiede, Anderson,

& Theriault, 2003; Thiede & Dunlosky, 1999; note that Dunlosky & Thiede, 2004, previously proponents of this position, have changed their views.) DR also posits that directing study to the most difficult items is the most effective study strategy, and a negative correlation between JOLs and study-time allocation, which this model predicts, has been taken to indicate an effective study-time allocation strategy. Focusing study where one is least competent seems intuitive (see Woodworth, 1921), and many studies have shown that people preferentially study the most difficult items (for a review, see Son & Metcalfe, 2000). These results provide support for DR, with one caveat: The experiments were all limited to situations in which it made sense for participants to try to master all of the materials because the time allowed for study was unlimited, so studying one item for longer did not mean another would be given short shrift, and because the items were all within a limited range of difficulty. It stands to reason that, to achieve mastery, the most difficult items must be studied most.

But total mastery is not always possible. In real life, time constraints often exist that prevent the learner from spending enough time on difficult items to master them. Under time pressure or when the performance goals are lower than complete mastery (see Dunlosky & Thiede, 2004), it may be suboptimal to emphasize items that are too difficult to learn. Indeed, T. O. Nelson and Leonesio (1988) have demonstrated that devoting excessive time to the most difficult items may not help learning, an effect that they dubbed the “labor-in-vain” effect. DR breaks down when people are given limited time or an easily attainable performance goal, when choosing one item means sacrificing another, and when people are faced with selecting only some of the items they do not know (Metcalfe & Kornell, 2005). When the only factor under consideration is choice of which items to study, as is the case in all of the experiments in the present article, then even proponents of DR (Dunlosky & Thiede, 1998, 2004), as well as those who argue that it is fatally flawed (e.g., Metcalfe & Kornell, 2005), have proposed alternative models.

According to an alternative, the region of proximal learning (RPL) framework (Metcalfe, 2002; Metcalfe & Kornell, 2003,

Nate Kornell and Janet Metcalfe, Department of Psychology, Columbia University.

This article is based in part on Nate Kornell's doctoral dissertation. Experiments 1, 2a, 2b, 2d, and 3a present new data that have never been published before but that were collected in conjunction with Experiments 1–4 and 6, respectively, from Metcalfe and Kornell (2005). This research was supported by National Institute of Mental Health Grant RO1-MH60637. We thank Lisa Son, Herb Terrace, Rob Hampton, Jim Magnuson, and Bridgid Finn for help and comments.

Correspondence concerning this article should be addressed to Nate Kornell, who is now at Department of Psychology, University of California, 1285 Franz Hall, Box 951563, Los Angeles, CA 90095. E-mail: nkornell@psych.ucla.edu

2005), people focus on studying the easiest items they do not know, prioritizing easier unknown items over the most difficult, especially when all of the unknown items cannot be studied. Prioritizing moderately difficult items is the most effective strategy according to RPL. When it is possible to try to master the entire set of to-be-learned items, RPL and DR make similar predictions because learning the difficult items will require the most study time. Indeed, DR reduces to a specific case of RPL. The difference occurs when people perceive that they have little or no chance of learning the most difficult items, given the constraints of the situation. Under such conditions, they will stop spending time on the most difficult items, instead turning to easier items that they believe they can learn.

In addition to predicting what people should and do study, RPL also identifies the processes by which study decisions are made. According to RPL, study-time allocation has two separable components: choice of items to study and perseverance on items once they are selected (Metcalf & Kornell, 2005). Study choices are made by selecting the easiest items one does not yet know and, among them, prioritizing those that are closest to being learned. Once a choice has been made, persistence depends on the perceived rate of information uptake, a measure that Metcalfe and Kornell called a judgment of the *rate* of learning (jROL). People are posited to persist until they perceive that they have ceased learning.

The evidence suggests that, when people choose what to study, they eliminate what they already know (e.g., Cull & Zechmeister, 1994; Masur, McIntyre, & Flavell, 1973; Son, 2004). But what happens next, when people chose among items they do not know (and when DR and RPL make opposite predictions)? Because almost all study-time allocation studies show that people select difficult items, one might expect people to select the most difficult of the unknown items, contradicting RPL. However, the data show no such contradiction. Metcalfe and Kornell (2005) reviewed the extant literature on study choice and concluded that “among the choice studies we have been able to find, even those showing very large negative gamma correlations, none have demonstrated that people preferentially chose the least learned items among those that were not already learned” (p. 470). Instead, the negative correlations appear to be due to the simple fact that, during the first phase of choice, people do not choose to study what they already know.

Since Son and Metcalfe’s (2000) literature review, which showed what was at the time a pervasive negative correlation between study-time allocation and JOLs, new evidence has come to light supporting RPL. Son and Metcalfe’s own data showed that, when total time studying was limited and thus more time studying one item meant less was available for others, people focused on easy items, supporting RPL and contradicting DR. Thiede and Dunlosky (1999) found a similar effect. Metcalfe (2002) and Metcalfe and Kornell (2003) replicated this finding using a different paradigm and showed that people focused on moderately difficult items. These latter studies also showed that, when more time was allowed for study, people increasingly focused on difficult items, which RPL predicts for two reasons. First, as time passes and learning progresses, the set of almost-but-not-quite-learned items (i.e., items in their region of proximal learning) shifts to more difficult items. Second, as study time increases, the possibility of mastery increases, and people become more inclined to

try to learn the most difficult items. Similarly, Metcalfe also found, using Spanish vocabulary pairs, that Spanish speakers studied more difficult materials than did non-Spanish speakers, although not the most difficult, again zeroing in on their own region of proximal learning. Finally, people shift to studying easier items when given goals that do not require learning all items (Dunlosky & Thiede, 2004; Thiede & Dunlosky, 1999).

These studies all suggest that assessing the goodness of metacognitive control in terms of the strength of the negative correlation between metacognition and study choice may be inappropriate. Instead, people seem to choose to study items that are in their region of proximal learning. However, not one of these studies has demonstrated that this strategy is effective. Indeed, despite its importance, the evidence for an effective monitoring–study–learning link is tenuous (T. O. Nelson et al., 1994; Thiede, 1999; Thiede et al., 2003).

Previous Findings on the Effectiveness of Metacognitively Guided Study

Making a link between metacognition and learning requires two skills on the part of the learner: monitoring learning and memory, and controlling study effectively based on that monitoring (T. O. Nelson & Narens, 1990). Thus, effective study involves linking three elements: monitoring, self-regulated study, and learning. Evidence for this link is far from conclusive.

One general approach to the monitoring–study–learning link is to attempt to show that people who are good at metacognitive monitoring, control, or both are also good learners. No consistent result has emerged from this area of research. Some studies have shown a positive correlation between metacognitive ability and learning ability (Bisanz, Vesonder, & Voss, 1978; Dunlosky & Connor, 1997; Maki & Berry, 1984; Schneider & Pressley, 1997; Yan, 1994). Others studies have failed to show such a link (Cavanaugh & Perlmutter, 1982; Cull & Zechmeister, 1994; Kearney & Zechmeister, 1989; Kelly, Scholnick, Travers, & Johnson, 1976; Lovelace, 1984; Maki & Swett, 1987; Moulin, Perfect, & Jones, 2000; Underwood, 1966). Thiede (1999) showed that test performance was positively correlated with both overall monitoring accuracy and study choice, but found no difference in monitoring accuracy or study choice between high- and low-memory performance groups (the later being an instance of “the learning ability paradox”; Cull & Zechmeister, 1994), demonstrating the inconsistency of the results within a single study. Furthermore, these studies are correlational and thus cannot determine causality; for example, both metacognition and learning may be two high-level cognitive skills that co-occur, but are not causally related (see Hager & Hasselhorn, 1992; Hasselhorn & Hager, 1989).

A second, more direct general approach to the monitoring–study–learning link has been to manipulate metacognitive monitoring, control, or both with the idea that, if people use these skills effectively, they should enhance learning. In general, when participants have not been allowed to control their study, the results have been negative or unclear. Begg, Martin, and Needham (1992) concluded that “memory monitoring does not make a valuable contribution to memory” (p. 212), and Hager and Hasselhorn (1992) found a similar negative result. Dunlosky, Kubat-Silman, and Hertzog (2003) found a positive effect, showing that a group of older participants trained in metacognitive monitoring outper-

formed other groups. This effect may have been due, however, at least in part, to the fact that the monitoring instructions recommended using self-tests to control learning, and self-testing alone has been shown to improve memory (Bjork, 1988; Kimball & Metcalfe, 2003; Roediger & Karpicke, 2006; Spellman & Bjork, 1992). Thus, further research with this promising technique is needed.

When people have been allowed to control their study and their metacognitive monitoring accuracy has been manipulated, the results have been ambiguous. In two such studies, which used different paradigms, materials, and participant groups; participants made either immediate or delayed JOLs (see Dunlosky & Nelson, 1992), and the delayed JOL condition was associated with better metacognitive monitoring and enhanced learning (Kennedy, Carney, & Peters, 2003; Thiede et al., 2003). This suggests that better monitoring may have led to better self-regulated study. However, Kimball and Metcalfe (2003) have shown that there is a memory-related confound in the delayed JOL paradigm, a kind of Heisenberg effect, whereby the act of making the judgment in the delayed JOL condition changes the memory itself. The difference in JOL accuracy disappeared when they removed this confound. These findings cloud the interpretation of Kennedy et al.'s results and may also be problematic for Thiede et al.'s interpretation of their data.

The studies most relevant to the predictions of DR and RPL compared self-regulated study with situations in which participants did not control their study. The first such study was by Atkinson (1972). He compared a condition in which people controlled their study with three computer-controlled conditions. Self-study resulted in fairly good learning, better than when the computer randomly assigned study time, but not nearly as good as the best computer-controlled condition. In the best condition, the computer focused time on the moderately difficult items, which RPL would predict to be the best strategy. In fact, Atkinson's model helped inspire RPL: He wrote that the most effective study was on items that were almost but not quite learned (Atkinson, 1972; Calfee & Atkinson, 1965). The fact that self-controlled learning was suboptimal may have been a result of the instructions participants were given, which emphasized mastery of all items (see Metcalfe, 2002). More recent findings suggest that people select in a manner that is very much in keeping with Atkinson's theory, if not his data (Metcalfe, 2000; Metcalfe & Kornell, 2005). Like Atkinson, Mazzoni and Cornoldi (1993, Experiment 3) found that self-controlled study was more effective than random study-time allocation.

T. O. Nelson et al. (1994) purported to have shown a positive metacognition–study–test link by manipulating whether or not participants controlled what they studied. After first studying 36 Swahili–English translations, college student participants were allowed to restudy half of the to-be-learned items in one of four conditions: One group was allowed to select the items they studied; one studied the subjectively easiest items (i.e., the items to which they had given the highest JOLs); one studied the subjectively most difficult items; and one studied the normatively most difficult items (based on norms from T. O. Nelson & Dunlosky, 1994). Memory performance was best for the self-selection and subjectively most difficult items groups. Both groups' performance was better than the easiest and normatively easiest groups' performance. The authors concluded that "this demonstrates that people can use their metacognitions to allocate their restudy ef-

fectively" (T. O. Nelson et al., 1994, p. 209). The results were also taken as support for a DR-like mechanism.

A crucial limitation of T. O. Nelson et al.'s (1994) study is that people were not forced to decide between moderately and highly difficult items—they could simply study all of the items they did not know. On the first test, before they made their choices about what to study, proportion correct was just under 50%. Thus, participants could simply choose to study the other 50%, that is, the items they had just demonstrated that they did not know. Both DR and RPL predict that this should be effective. Note that Atkinson's (1972) participants had already demonstrated that they could learn more than with random choice of items by selecting the unknown items, but that it was *not* the optimal strategy. In this light, a test of optimality becomes more interesting when people are forced to prioritize which unlearned items they want to study, and that is also when DR and RPL make different predictions (Metcalfe & Kornell, 2005).

The Present Experiments

In summary, the causal link between metacognition, study choice, and enhanced learning is unresolved. The tack we took in testing this link was to compare a self-regulated study condition with control conditions that reversed or otherwise violated the participant's study choices. The logic is, if people do better when allowed to control their study than when in the control condition, their study strategy must be effective. Obviously, the more effective the control condition, the more impressive a positive finding will be. Random allocation of study time is one control condition that has been used before; another, used in the present experiments, is to allow participants to study the opposite of what they requested. In both cases, if self-guided study results in better learning than the control condition, it is the result of an effective study strategy. In some cases, other conditions that should lead to very effective learning can be devised. For instance, if DR is correct, then allocating study selectively to the most difficult items should be the optimal strategy and result in the most learning. If RPL is correct, then allowing study of the easiest as-yet-unlearned items should be effective. These other conditions were also used to investigate the degree to which, given their own choices, people are able to optimize study.

Accordingly, the first experiment presented here tested whether metacognitively guided study positively affects learning in a semantic memory task. Experiments 2a–d also examined the effectiveness of metacognitively guided study. Two variables were manipulated: the degree of learning of the items and whether the computer or the participants controlled study. Experiments 3a and 3b measured the effectiveness of metacognitively guided study in a situation in which the already-known items were eliminated and when DR and RPL make opposite predictions about what people will study. This has clear theoretical implications for current models of study-time allocation. It is also a realistic situation with potentially important pedagogical implications. Experiments 3a and 3b also included computer-controlled conditions to address the question of which model makes the correct prediction about how people should study by comparing people's performance with the best computer-controlled conditions.

Experiment 1

The overall logic of this experiment was the same as the experiments that follow. The procedure required participants to choose which items they wanted to study. The independent variable was whether their study choices were honored or not: Participants were allowed to study what they chose in the honor condition or the opposite in the dishonor condition. In this first experiment, there was also an intermediate condition in which participants studied half of the items they chose and half of the items they did not (the 50/50 condition). During the study choice, participants were allowed to choose half of the to-be-learned items, in an effort to make them choose carefully, instead of simply choosing all of the items.

Method

Participants, design, and materials. Twenty-four Columbia University undergraduates participated for course credit or pay. They were treated in accordance with the ethical principles of the American Psychological Association.

The design was a one-way within-participants design with three within-participants conditions: honor, 50/50, and dishonor (which will be described below).

The materials were 200 general information questions that were selected from the norms published by T. O. Nelson and Narens (1980). For example, one question was, *What is the last name of the man who supposedly assassinated John F. Kennedy?* (Oswald). This and the experiments that follow were conducted on Macintosh computers.

Procedure. The experiment consisted of two phases, choice and study-test. During the choice phase, 192 trivia questions were presented individually, along with two buttons labeled *Study* and *Don't study*. The order of the items was randomized for each participant. Participants were told to press *Study* if they wanted to see the question and answer together later (during the study-test phase) and to press *Don't study* if they did not. They were also told that they were allowed to select only half of the items for further study. To make that more manageable, instead of requiring people to request 96 of the 192 items, the items were randomly split into six 32-question sets during the choice phase. A counter below each choice button displayed the number of times it had been pressed as well as the number of times it could be pressed; for example, if *Study* had been pressed four times, its counter would read 4/16. When one button had been pressed 16 times, it became unpressable. When both reached 16, the counters reset to 0, and a new set of 32 questions began. To limit the time participants took during the choice phase, the computer played recorded voices saying "hurry!" after 8 s and "please choose now!" after 16 s. The "hurry" prompt was used during the choice phase so that participants did not spend time trying to recall or encode the answer when they were supposed to be choosing. It was meant to allow enough time so that participants did not feel rushed.

At the end of the choice phase, each of the six sets of 32 questions was assigned to one of the three conditions (with two sets per condition). The order of conditions was assigned randomly for each participant. In the honor condition, the 16 requested items were presented for study; in the 50/50 condition, eight requested and eight nonrequested items were randomly selected and presented; and in the dishonor condition, the 16 nonrequested items were presented. During the initial instructions, participants were told that the items they were allowed to study would not always match the items they had requested.

During the second phase, the study-test phase, words were presented for study and then tested, one set of words at a time. During study, the 16 questions designated for study were presented individually, with their answers, for 5 s each. Following study, there was a distractor task, which required participants to count backward from a random three-digit number by threes for 15 s, and then all 32 questions from that set were tested in

random order. During the test, the questions were presented individually, and participants were asked to type in the answer and to press *return*. Again, to limit the time participants took during the test, after 12 s, a recorded voice said, "Hurry!" After 24 s, the computer went on to the next question automatically, and a recorded voice said, "Next question." These responses were considered incorrect. Participants were informed that they would be given a maximum of 24 s to answer. The "hurry" prompts were used during the test to limit differences in timing across participants.

Before beginning the experiment, participants were read instructions and then completed a practice list with eight questions, which were assigned to the honor condition. Following the practice list, participants were given a chance to ask the experimenter questions, and then the experiment began. The practice list was included in this experiment (and those that follow) to ensure that people understood the task before beginning the experimental session. This was crucial because we thought participants might choose their study strategies differently depending on how clearly they understood the task. It was also used to establish that they would be allowed to study what they chose, but tested on all items. The practice list was not included in the analyses.

Results and Discussion

In analyzing this and the experiments that follow, forced-choice trials were not removed from the analyses. Forced-choice trials occurred when participants exhausted one study option and the other option became the only possible choice (e.g., once the participant chose *Study* 16 times, the remaining choice had to be *Don't study*). Removing forced-choice trials seems reasonable because they do not necessarily reflect what participants want to study (and it would be appropriate if we were analyzing only study choices). However, doing so would be inappropriate. To illustrate, imagine a participant who uses up the 16 study choices in the first 18 trials. (Participants tended to exhaust the *Study* option first.) What happens if only those 18 items are included in the analysis? In the honor condition, the participant will have been allowed to study 16 of the 18 items included in the analysis; in the dishonor condition, the participant will have been allowed to study only 2 of those items, and the other 14 items studied will have been excluded. This is unacceptable because it will lead to better performance in the honor condition regardless of study strategy. Thus, it is necessary to include all items. Doing so is the more conservative approach because it adds only noise to the data. The criterion for significance was set at $p < .05$ throughout. Effect sizes (ES s) were computed using partial η^2 .

Test accuracy was best in the honor condition ($M = .47$), followed by the 50/50 condition ($M = .44$) and then the dishonor condition ($M = .38$) (see Figure 1). The main effect of condition was significant, $F(2, 23) = 16.22, p < .0001$ ($ES = .41$). Tukey's post hoc tests showed that scores in the 50/50 and honor conditions were higher than those in the dishonor condition, but that scores in the honor and 50/50 conditions did not significantly differ. Performance in the 50/50 condition was also analyzed separately, comparing performance for the honored and dishonored pairs. Accuracy was significantly higher for the honored pairs ($M = .49$) than for the dishonored pairs ($M = .40$), $t(23) = 3.79, p < .001$ ($ES = .35$). Thus, this experiment demonstrates that people can make adaptive study decisions, which lead to more effective learning.

The participants' study strategy was to choose to study the items they did not know. An analysis of question difficulty based on T. O. Nelson and Narens' (1980) norms showed that items selected for study were significantly more difficult than those not selected,

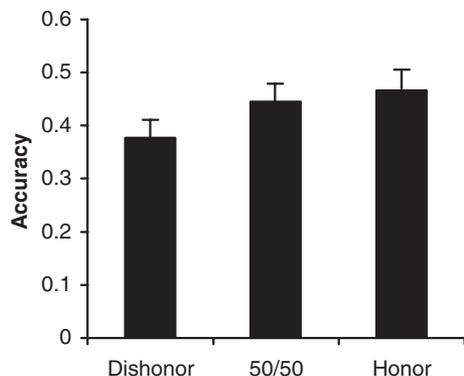


Figure 1. Proportion correct in Experiment 1. Error bars represent standard errors.

$t(23) = 7.88, p < .0001$ ($ES = .73$). This was confirmed by an analysis of Goodman–Kruskal’s gamma correlations, a technique commonly used in the analysis of metacognitive accuracy (T. O. Nelson, 1984). The mean gamma correlation between study choice and item difficulty was significantly greater than zero ($M = .28$), $t(23) = 7.82, p < .0001$. In addition, self-reports showed that 20 of the 24 participants said that their strategy was to study the items they did not know and that none said the opposite. In summary, Experiment 1 showed that the strategy people used was to study the items they did not know and that this proved to be effective. The next experiments were designed to replicate and extend this finding.

Experiments 2a–d

Experiments 2a–d were designed to determine whether people would make effective study choices in a situation in which new learning was required. In these experiments, study control and degree of learning were manipulated independently.

In each of these experiments, the procedure began with an initial study phase in which only half of the pairs were presented for study, leaving the other half unstudied and unknown. The items that were presented were shown either one or six times each. The participants were then asked to select half of the items for study (from among all items), and they were allowed to study again in either the honor or dishonor condition. Finally, they were tested. Experiments 2a–d were all variations on this procedure. The differences will be described as each experiment is introduced.

The reason for varying the number of presentations was to manipulate learning. Choosing new items over items that were already presented, according to RPL, should be effective only when the presented items have been well learned. If the presented items have not been learned but are already familiar, then they should be the most likely items to be in the person’s region of proximal learning and should be chosen instead of completely new items. Thus, in the six-presentation condition, in which the items were assumed to be well learned, the optimal choice should be to study the new, unrepresented items. In the one-presentation condition, however, there should be less of a tendency to choose the completely new items because the as-yet-unlearned but previously presented items should be likely to benefit from further study, making them good candidates for choice.

The questions were, first, would people choose to study the items that had not been presented; second, would they be more likely to do so in the six-presentation condition than in the one-presentation condition; and third and most important, would allowing participants to study their own choice of items lead to improved performance.

Experiment 2a

Method

Participants, design, and materials. Twenty-four Columbia University undergraduates participated for course credit or pay. They were treated in accordance with the ethical principles of the American Psychological Association. The design was a 2 (honor vs. dishonor condition) \times 2 (one vs. six study trials) within-participants design, with the order of presentation of these conditions counterbalanced across participants.

The materials were 104 word pairs selected on the basis of cue-to-target associability (D. L. Nelson, McEvoy, & Schreiber, 1998). All pairs fell within the .050–.054 range, for example, *sickle–hammer* and *well–done*. These pairs were selected to be somewhat learnable in a single exposure, but not guessable without any exposure.

Procedure. In this experiment, four separate lists of word pairs were presented and tested. Each list was assigned to a different condition (e.g., the honor, one-presentation condition). Unlike in Experiment 1, in which each list was presented in both of the two phases, in this experiment, each list was completed before the next list began. The procedure for each list included four phases: study, choice, restudy, and test. There were 24 word pairs per list.

During the first phase, study, 12 of the 24 word pairs were presented for study. The computer randomly decided which pairs to present. The reason for presenting only half of the pairs was so that the other half would be guaranteed to be unknown. In the one-presentation condition, the 12 pairs were presented once each in random order; in the six-presentation condition, the 12 pairs were presented six times, and the order of presentation was reshuffled randomly each time. Each presentation lasted 2.5 s. At the end of study, there was a distractor task, which required participants to count backward by threes for 15 s.

The second phase, choice, was similar to the choice phase in Experiment 1: Each of the 24 cues was shown individually, and participants were allowed to choose to study half of them. There were counters indicating how many times each option had been selected (e.g., at the beginning, the counters under both the *Study* and *Don’t Study* buttons read 0/12). The computer played a recorded voice saying “hurry!” after 8 s and “please choose now!” after 16 s.

During the third phase, restudy, participants were shown half of the word pairs to study for 4 s each. There were two conditions: honor, in which the 12 requested items were presented, and dishonor, in which the 12 nonrequested items were presented. (*Restudy* is actually something of a misnomer because some of these pairs had not necessarily been studied yet.) At the end of the third phase, there was another 15-s distractor task.

Finally, during the test, each of the 24 cue words was presented individually, and the participant was asked to type in the target and press *return*. After 12 s, a recorded voice said, “Hurry!” After 24 s, the computer automatically went on to the next question, and a recorded voice said, “Next question.”

When a list ended, the next list began. With four conditions and four lists, 24 orders of conditions are possible (e.g., H1, D6, D1, H6). Each participant was assigned to 1 of the 24 orders of conditions randomly such that no 2 participants were assigned to the same order.

To begin the experiment, participants were read instructions, and then they completed a practice list. The practice list included eight pairs and was assigned to the one-presentation honor condition.

At the end of the last list, the computer asked participants to type the answers to three questions. (a) *How did you decide which items you wanted*

to study? (b) Did you find it easy to concentrate during study? Or did you get bored? And did this change as time went on? (c) When you were making your choices about what to study, did you ever choose the opposite of what you actually wanted to study (maybe because you thought the computer would not honor your requests anyway)? If so, do you remember which list(s)? Participants who answered yes to the last question were replaced. This resulted in 6 participants being replaced.

Results and Discussion

As predicted, test accuracy in the honor condition was significantly better than in the dishonor condition, $F(1, 23) = 16.20, p < .001$ ($ES = .41$), indicating that people benefited from being allowed to control their own study (see Figure 2). Not surprisingly, performance was also better in the six-presentation condition than in the one-presentation condition, $F(1, 23) = 14.31, p < .01$ ($ES = .38$). Contrary to the hypothesis that the effect of honoring choices would be larger in the six-presentation condition than in the one-presentation condition, the interaction was not significant, $F(1, 23) = 1.65, p = .21$ ($ES = .07$).

What study strategy did participants use? To answer this, gamma correlations between study choice and whether or not an item had been presented were computed. A negative correlation indicates that participants chose items that had not been presented. The average gamma correlation was significantly negative in both the one-presentation condition ($M = -.62$), $t(23) = -4.95, p < .0001$, and the six-presentation condition ($M = -.76$), $t(23) = -6.35, p < .0001$. Comparing the two conditions, the preference for unfamiliar items was stronger in the six-presentation condition than in the one-presentation condition, but the difference was only marginally significant, $t(23) = 1.77, p < .10$ ($ES = .12$). This finding, that study choices were not significantly different in the one- and six-presentation conditions, is consistent with the idea that participants were able to learn the presented items even in the one-presentation condition and consequently chose not to study them.

Participant self-reports confirmed the preference for unknown items: 20 of the 24 participants reported selecting the items they did not know. The data from the 4 participants who did not report studying the unknown items showed that they were the least likely to select unfamiliar items for study, as they reported. They accounted for the four lowest gamma correlations, two of which were

highly positive (meaning they chose to study the items they had already seen). Because these outliers were included in the mean gamma correlations between study choice and familiarity ($-.62$ and $-.76$ for the one- and six-presentation conditions, respectively, as noted above), the median gamma correlations ($-.87$ and $-.99$, respectively) were even more extreme than the means. The 4 participants who did not report studying unknown items also did worse in the memory test as a result. Being allowed to control their own study was less beneficial than it was for other participants. To measure study benefit, the difference in performance between the honor and dishonor conditions was computed, because the better the study strategy, the more honoring it should improve performance. These 4 participants had the four lowest honor–dishonor difference scores, three of which were negative (meaning they did worse when given the items for study that they requested).

To summarize, the results demonstrated, first, that people chose to study the items they did not know, and second, that this was an effective strategy: People benefited from being allowed to control how they studied. Study control was not significantly more beneficial in the six-presentation condition than in the one-presentation condition, presumably because the materials were highly associated and learnable, even with a single presentation, and the distinction between learned and unlearned items was thus clear in both conditions.

Experiment 2b

Experiment 2b was designed to replicate Experiment 2a, but with more difficult materials. We made the task more difficult in an effort to decrease learning in the one-presentation condition while leaving the six-presentation condition relatively intact. The goal was to cause some of the presented items in the one-presentation condition to fall into the region of proximal learning, making them closer to being learned than items that were completely new but still as yet not well learned. Such items were expected to be chosen in preference to new items.

Method

The only difference between Experiments 2a and 2b was the stimuli. Whereas the stimuli in Experiment 2a were pairs of semantically related

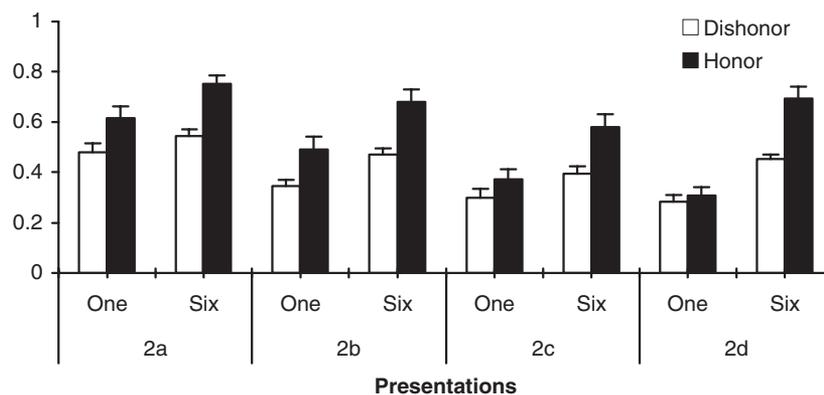


Figure 2. Proportion correct in Experiments 2a–d. Error bars represent standard errors.

words, the stimuli in Experiment 2b (and in Experiments 2c and 2d) were unrelated word pairs (e.g., *troops-coffee*). The words were selected from norms published by Paivio, Yuille, and Madigan (1968). Two-hundred eight words were selected that scored 6 or 7 on a 1–7-point scale of both concreteness and imagery. Words that were synonyms of other words on the list were excluded. These words were randomly paired to make 104 pairs. The participants were 30 Columbia University undergraduates. Six participants had to be removed because they reported choosing the opposite of what they wanted to study based on the expectation that their choices would not be honored, leaving 24 participants.

Results and Discussion

The results of Experiment 2b closely matched those of Experiment 2a. Performance was significantly better in the honor condition than in the dishonor condition, $F(1, 23) = 24.65, p < .0001$ ($ES = .52$; see Figure 2). Performance was also better with six presentations than with one, $F(1, 23) = 38.87, p < .0001$ ($ES = .63$). Once again, honoring choices was not significantly more beneficial in the six-presentation condition than in the one-presentation condition, $F(1, 23) = 1.26, p = .27$ ($ES = .05$).

In terms of study strategy, gamma correlations showed that participants tended to select previously unrepresented items in both the one-presentation condition ($M = -.54$), $t(23) = -5.31, p < .0001$, and the six-presentation condition ($M = -.73$), $t(23) = -5.65, p < .0001$. Comparing the two conditions showed that the gamma correlation was stronger in the six-presentation condition than in the one-presentation condition, $t(23) = 1.96, p < .10$ ($ES = .14$), but the difference was only marginally significant.

Self-reports again confirmed the preference for unrepresented items: 21 of the 24 participants reported selecting items they did not know. The 3 remaining participants had the three lowest study–familiarity gamma correlations, all of which were positive (indicating they chose to study the items they had seen before). Again, as a result of these outliers, the median gamma correlations ($-.71$ and $-.996$ for the one- and six-presentation conditions, respectively) were more extreme than the means ($-.54$ and $-.73$, respectively). These 3 participants had three of the four lowest honor–dishonor difference scores, and all three scores were negative (meaning the participants did worse when allowed to study the items they chose). Thus, whereas most participants selected the items they did not know, some did not, and they performed poorly as a consequence.

The results of this experiment confirmed the findings from Experiment 2a: People studied adaptively, and the strategy they used was to choose the unknown items, but the expected interaction between honoring choices and amount of learning was still not statistically reliable.

Experiment 2c

Experiment 2c was designed with the same goal as Experiment 2b. The task was made more difficult to further decrease learning.

Method

Experiment 2c was identical to Experiment 2b with one exception. In Experiments 2a and 2b, each word pair was presented for 2.5 s during the initial study phase. In Experiment 2c, the time was shortened to 1 s to make the task more difficult. The participants were 30 Columbia University undergraduates. Six participants had to be replaced because they reported

choosing the opposite of what they wanted to study based on the expectation that their choices would not be honored, leaving 24 participants.

Results and Discussion

As in the previous experiments, performance was better in the honor condition than in the dishonor condition, $F(1, 23) = 30.93, p < .0001$ ($ES = .57$; see Figure 2), and with six presentations than with one, $F(1, 23) = 46.08, p < .0001$ ($ES = .67$). There was a significant interaction, $F(1, 23) = 5.68, p < .05$ ($ES = .20$), showing that having one's study choices honored was more beneficial in the six-presentation condition than in the one-presentation condition.

The mean gamma correlations between study choice and whether or not an item had been presented were significantly below zero in both the one-presentation condition ($M = -.22$), $t(23) = -3.15, p < .01$, and the six-presentation condition ($M = -.61$), $t(23) = -6.23, p < .0001$, although, as predicted, the negative gamma correlation was significantly stronger for the six-presentation condition than for the one-presentation condition, $t(23) = 3.42, p < .01$ ($ES = .34$).

Self-reports indicated that 19 of the 24 participants chose to study the items they said they did not know. The 5 remaining participants accounted for the five highest choice gamma correlations, three of which were positive and one of which was zero. Once again, as a result of these outliers, the median gamma correlations ($-.32$ and $-.80$ for the one- and six-presentation conditions, respectively) were more extreme than the means ($-.22$ and $-.61$, respectively). These participants' performance also suffered: All five had honor–dishonor difference scores below the median, including the two lowest scores.

In summary, people chose to study the unstudied items, especially in the six-presentation condition, and benefited from being allowed to control how they studied. Moreover, in this experiment, the preference was more pronounced when the difference between what was learned and not learned was more distinct, that is, in the six-presentation condition rather than in the one-presentation condition. Study control was also more beneficial in the six-presentation condition than in the one-presentation condition. These findings fit with the RPL prediction that people should choose to study items they have been exposed to but have not learned in preference to items they have no knowledge of.

Experiment 2d

Experiment 2d was designed to replicate and confirm the results of Experiment 2c while increasing the difference in learning between the one- and six-presentation conditions.

Method

The difference between Experiments 2c and 2d was that, in the one-presentation condition, the pairs were presented for only .5 s each, whereas in the six-presentation condition, they were presented for 2.5 s each. The participants were 25 Columbia University undergraduates. One participant was replaced after he or she reported choosing the opposite of what he or she wanted to study based on the expectation that his or her choice would not be honored, leaving 24 participants.

Results and Discussion

The results of Experiment 2d were very similar to those of Experiment 2c. Performance was better in the honor condition than

in the dishonor condition, $F(1, 23) = 23.39, p < .0001$ ($ES = .50$; see Figure 2), and with six presentations than with one, $F(1, 23) = 184.06, p < .0001$ ($ES = .89$). The interaction was also significant, $F(1, 23) = 38.57, p < .0001$ ($ES = .63$), showing that honoring study choices was more beneficial in the six-presentation condition than in the one-presentation condition.

The gamma correlation between study choice and whether or not an item was presented was significantly below zero in the six-presentation condition ($M = -.93$), $t(23) = -22.11, p < .0001$, showing that participants chose to study the previously unstudied items in that condition. By contrast, the gamma correlation was not different from zero in the one-presentation condition ($M = -.04$), $t(23) = -.69, p = .50$. The gamma correlation was significantly more negative in the six-presentation condition than in the one-presentation condition, $t(23) = 12.26, p < .0001$ ($ES = .87$).

It appears that, with the very short presentations, people frequently decided to study already-presented items again. RPL predicts that participants should choose to study previously presented items if they have not yet learned them, and may do so even in preference to completely new items, because these items are closer to being learned than completely new items. Doing so exclusively would lead to a positive correlation between choice and whether or not an item is presented, however. The fact that the mean correlation was approximately zero in the one-presentation condition could indicate that people were using a study-choice strategy, as suggested by RPL. Alternatively, the zero correlations might indicate that participants were simply choosing randomly.

According to self-reports, 20 of the 24 participants selected the items they did not know. The 4 remaining participants accounted for four of the five weakest gamma correlations. Because of these outliers, the median gamma correlation in the six-presentation condition ($-.996$) was more extreme than the mean ($-.93$). In the one-presentation condition, the median was zero, approximately the same as the mean ($-.04$). In terms of how much they benefited from their study choices, the 4 deviant participants did not do as poorly as in previous experiments. In terms of the honor–dishonor scores, two were below the median, with one score being negative, but the other two were above the median. This may have been a result of the fact that, at times, studying previously presented items was actually a good strategy.

In summary, people chose to study the items they did not know, and being allowed to control how they studied improved their performance. The results in the one-presentation condition indicate that people may have been zeroing in on almost-learned items, as suggested by RPL, but they are equally supportive of a simple random choice.

Discussion

In Experiments 2a–d, people benefited from being allowed to control how they studied, performing better in the honor condition than in the dishonor condition across the board. In Experiments 2c and 2d, gamma correlations between study choice and whether or not the items had been presented were higher in the six-presentation condition than in the one-presentation condition. In these experiments, having one's study choices honored was more beneficial in the six-presentation condition than in the one-presentation condition. Neither of these effects was significant

when participants had more time to learn the items in the one-presentation condition in Experiments 2a and 2b. RPL predicts that, when some items are well learned, people should unambivalently choose to study the items that are as yet completely unlearned, as was shown in all of the six-presentation conditions in this series. However, when the items are slightly learned but not yet mastered, as in the one-presentation conditions of Experiments 2c and 2d, they should be in the person's RPL. These items may be chosen for study even in preference to completely unlearned items because they are the most likely to yield learning gains in a small amount of additional study time. The shift in gamma correlations in the one-presentation condition toward zero across experiments is consistent with this explanation, but, because it is also consistent with participants' simply choosing randomly, the results require further clarification. Thus, we conducted Experiments 3a and 3b.

Experiments 3a and 3b

Experiments 3a and 3b were designed with two main goals. The first was to test whether people benefited from controlling how they studied in a situation in which they could not simply choose to study anything they did not already know. Instead, they were asked to make a more nuanced decision, which was accomplished by asking participants to choose to study half of a set of items, all of which they had already demonstrated they did not know. This allows a crucial comparison of two models, DR and RPL. DR predicts that people will choose to study the most difficult items and that to do otherwise would hurt learning. RPL predicts the opposite, that people will choose the easiest of the items they have not yet learned, that is, the items closest to being learned and deepest in the region of proximal learning. It also predicts that this is the best strategy.

The second goal was to try to address the question of how effectively participants control their study. The experiments presented so far show, in a variety of situations, that people make adaptive study choices, which positively impact their learning. However, the fact that a strategy works does not mean it could not work better. Atkinson (1972) found better performance when participants controlled their own study than when the computer gave them items randomly, but performance was clearly not optimal. The best computer-controlled condition led to much better performance than did the self-control condition. Experiments 3a and 3b were modeled on the methodological approach of Atkinson and T. O. Nelson et al. (1994). A condition in which people controlled their own study was contrasted not only with a condition in which their choices were dishonored, but also with two computer-controlled conditions, one that should lead to the most effective study according to DR (studying the subjectively most difficult items) and one that should optimize performance according to RPL (studying the subjectively easiest but as-yet-unlearned items). Comparison of these conditions distinguishes between the models and also provides evidence concerning how metacognitively adept participants are. Performance that is worse in the self-study condition than in the best computer-controlled condition indicates suboptimal study; performance that is as good as or better than the best computer strategy could be said to demonstrate a high degree of metacognitive acuity (although it will not prove that participants study optimally, of course, because it is impossible to

eliminate the speculation that performance could have been even better with some hypothetical optimal strategy that we did not utilize).

To summarize the procedure, participants were presented with Spanish–English translations of varying difficulty to study. After studying each pair, they made a JOL. An initial test on the items followed, and items answered correctly were excluded from the rest of the experiment. After the initial test, participants were asked to choose half of the items to study from among the remaining pool of unlearned items. After making their choices, participants restudied in one of four conditions: high JOL, in which they studied the highest JOL items; low JOL, in which they studied the lowest JOL items; honor, in which they studied what they chose; and dishonor, in which they studied what they did not choose. Finally, they were tested on all items (excluding those answered correctly on the initial test).

Experiment 3a

Method

Participants, design, and materials. Forty-eight Columbia University undergraduates participated for course credit or pay. They were treated in accordance with the ethical principles of the American Psychological Association.

The design was a one-way within-participants design with four restudy conditions: high JOL (the easiest items), low JOL (the most difficult items), honor, and dishonor.

The materials were 144 Spanish–English translations from Metcalfe (2002). Three levels of difficulty were used: The easy items were cognates (e.g., *family–familia*); the moderately difficult items had some discernible connection between the English and Spanish words (e.g., *husband–marido*); and the difficult items were very difficult (e.g., *sharpener–sacapuntas*).

Procedure. Experiment 3a consisted of four lists, each of which was completed before the next began. Each list included 24 items: 8 were easy, 8 were moderately difficult, and 8 were difficult. The procedure for each list consisted of five phases.

In the first phase, participants studied and made JOLs. Each word pair was presented for 4 s, after which it disappeared, and participants were asked to make a JOL. During the instructions, the JOL rating was defined as follows: *Rate how confident you are that in a few minutes you will be able to come up with the Spanish word when shown the English word.* The participants were shown six buttons labeled 0%, 20%, 40%, 60%, 80%, and 100%. Because the time allowed for the JOL could also serve as study time, if the participant had not responded after 4 s, the computer said the word “hurry” aloud, and if there was still no response after another 4 s, it said, “Please choose now.” At the end of the first phase, there was a distractor task during which participants counted backward by threes for 15 s.

The second phase was the initial test. Each English word was shown individually in random order, and the participant had to type in the Spanish translation and press *return*. The purpose of this phase was to identify and eliminate the items people already knew. When the participant responded correctly, the item was removed from the remainder of the experiment. Responses had to be spelled correctly to be considered correct. If an odd number of items remained at the end of the phase, an additional item was randomly selected and removed to make the number remaining even.

In the third phase, participants chose which of the items they wanted to restudy. The instructions made clear that they would be deciding among a set of items they had answered incorrectly. Each of the cues was shown individually, accompanied by buttons labeled *Study* and *Don't study*. Counters below the buttons tracked how many times each had been pressed

and showed the total number of times they could be pressed. The number of times each could be pressed varied; for example, if the participant answered four items correctly during the initial test, 20 would remain, and each counter would start at 0/10. If participants took more than 8 s to choose, the computer said, “Hurry!” After an additional 8 s, it said, “Please choose now.”

In the fourth phase, half of the items were presented for restudy for 4 s each in one of four conditions. In the honor condition, the requested items were presented; in the dishonor condition, the nonrequested items were presented. In the other two conditions, the study requests were ignored, and the items that had not been answered correctly in the initial test were median split into two groups based on JOLs. For the high-JOL condition, the items given the highest JOLs were presented; for the low-JOL condition, the items given the lowest JOLs were presented. Participants were told during the instructions that what they were allowed to study would not always match their choices. After all the items had been presented, there was a distractor task during which participants counted backward by threes for 15 s.

The final phase was the test. Each English word was shown individually, and the participant was asked to type in the Spanish translation and press *return*. If no answer had been given after 12 s, the computer said, “Hurry!” If there had still been no response after 24 s, the computer went on to the next question automatically and said, “Next question.” Participants were informed that they would have a maximum of 24 s to answer.

The experiment proper consisted of four lists. There was also an additional list at the start of the experiment that, unbeknown to the participants, was a practice list. The practice list was used to allow participants to become familiar with the procedure, the varying degrees of difficulty of the words, and the potential consequences of different study strategies on their test performance. This first list was not included in the analyses, and it was always assigned to the honor condition to establish the importance of the study choices. Following the practice list, there were four additional lists, which were each assigned to one of the four restudy conditions. The order of the conditions of the four lists was counterbalanced across participants, with 1 participant randomly assigned to each of the 24 possible orders of conditions.

The experiment began with instructions, which were spoken aloud by the computer in addition to being shown on the screen. The instructions alternated between explaining each phase of the experiment and allowing the participants to practice using a short list of 12 items.

After completing the last list, participants were asked the same three questions as in the previous experiments. Eleven of the 48 participants were replaced because they said they chose to study the opposite of what they really wanted with the expectation that they would not be allowed to study what they chose. Because this is a fairly large number of participants, it seemed possible that excluding them would significantly bias the results. Therefore, the key analyses of Experiments 3a and 3b, which are presented below, were repeated with all participants included. Excluding these participants did not bias the findings, and none of the results reported below changed when all participants were included in the analyses.

Results and Discussion

There are four main questions to be answered by analyzing the data. What did people choose? What should they have chosen? Did they do better when they controlled their own study? How adaptive were their study choices? These will be answered in order. But first, to confirm that participants had good metacognitions in this situation and made accurate JOLs, we computed a gamma correlation for each participant between the initial JOL and performance on the first test. The mean gamma correlation was significantly greater than zero ($M = .82$), $t(47) = 47.48$, $p < .0001$. (The reason that this correlation is stronger than the literature on immediate

JOLs would predict is probably because there were large differences in objective difficulty between the easy, moderately difficult, and difficult translations, making them relatively easy to discriminate.)

Participants choose to study the easiest as-yet-unlearned items. The mean gamma correlation between JOL and study choice was significantly greater than zero ($M = .32$, $t(47) = 5.21$, $p < .0001$). A comparison of JOL levels produced a similar result: The items participants chose to study were easier (mean JOL = 35.27) than those they chose not to study (mean JOL = 24.79), $t(47) = 4.92$, $p < .0001$ ($ES = .34$). These findings support RPL and contradict DR.

Self-reports also showed a preference for easy items: 37 of the 48 participants reported selecting the easiest items. The 11 remaining participants all had study–JOL gamma correlations well below the median, and they accounted for 9 of the 10 lowest gamma correlations, all 9 of which were negative (meaning they chose to study the items they found relatively difficult). As a result of these participants, the median gamma correlation (.42) was higher than the mean (.32). These participants also did worse as a result of their study strategy. They accounted for six of the eight lowest honor–dishonor difference scores, and they had significantly higher honor–dishonor difference scores compared with the other 37 participants, $t(46) = 3.12$, $p < .01$. Thus, most participants selected the easy items in this experiment, but a minority did not, and their performance suffered as a consequence.

The three remaining major questions all concern which conditions led to the best performance. There was a significant main effect of condition on test accuracy, $F(3, 47) = 25.58$, $p < .0001$ ($ES = .35$). In order from best to worst, performance was as follows: high JOL ($M = .28$), honor ($M = .22$), dishonor ($M = .13$), low JOL ($M = .12$; see Figure 3). Tukey's post hoc tests showed that all differences were significant except for the comparison of the dishonor and low-JOL conditions.

What should people have chosen to study? The theories point to a comparison of the high- and low-JOL conditions because, according to RPL and DR, respectively, these conditions should lead to optimal performance. As shown above, performance was significantly better in the high-JOL condition ($M = .28$) than in the low-JOL condition ($M = .12$), supporting RPL.

Participants benefited from controlling their own study. Performance was better in the honor condition ($M = .22$) than in the

dishonor condition ($M = .13$). Because the results above showed that people studied the easy items and that the best strategy was to study the easy items, this should come as no surprise. It is, however, the first demonstration of study control benefiting learning in a situation in which people could not simply choose every unknown item, but instead chose the easier items from among items they did not know, as predicted by RPL.

How adaptive were participants' study choices? This can be answered by comparing the honor condition ($M = .22$), in which participants were allowed to control their study, and the high-JOL condition ($M = .28$), which RPL predicts should lead to excellent performance and which led to the best computer-controlled performance in this experiment. These two conditions were significantly different, which indicates suboptimal study. This result is somewhat misleading, however, because it assumes that every participant studied effectively by focusing on the easy items, whereas the results above showed that there were 11 participants who did not report studying the easy items and did worse as a result. To determine how adaptive typical participants' study choices were requires excluding the 11 participants who did not report studying the easy items. After doing so, the main effect of condition was still significant, $F(3, 36) = 25.08$, $p < .0001$ ($ES = .34$), but the high-JOL and honor conditions were no longer significantly different. The accuracy of the four groups was as follows: high JOL ($M = .26$), honor ($M = .22$), dishonor ($M = .10$), and low JOL ($M = .11$). Tukey's post hoc tests showed that all differences were significant except for the comparisons of the dishonor and low-JOL conditions and the honor and high-JOL conditions. Thus, when only participants who chose to study easy items were included, the condition in which people controlled their own study was not significantly different from the best computer-controlled condition. This shows that performance in the honor condition was highly adaptive.

There is another reason why the honor condition might result in worse performance than the high-JOL condition: In the honor condition, some of the choices participants made were forced, because the other option had already been exhausted (e.g., they could not press *Study* anymore). These forced choices could lead to suboptimal study. As noted earlier, however, there is a problem with excluding forced choices from the analysis because it creates a bias in the data, whereby the majority of the items in the honor condition have been studied, which is not necessarily the case in the high-JOL condition. Nevertheless, including these forced choices might obscure somewhat the "real" performance under conditions of self-choice. We therefore conducted a final analysis. To make a fair comparison of the conditions, all forced-choice items were excluded from the analysis. We then went through each participant's data for each condition separately; whenever the number of items they had studied was greater than the number they had not studied or vice versa, the excess items were randomly excluded. As a result, there were equal numbers of studied and unstudied items in both conditions, none of which were forced. Again, the 11 participants who reported using deviant strategies were excluded from the analysis. An unavoidable drawback of this analysis is that an additional participant had to be excluded from the analysis because of a lack of observations in the honor condition, and many of the participants who were not excluded had very few observations. Nonetheless, test accuracy in the high-JOL condition ($M = .25$) was not significantly different from that in the

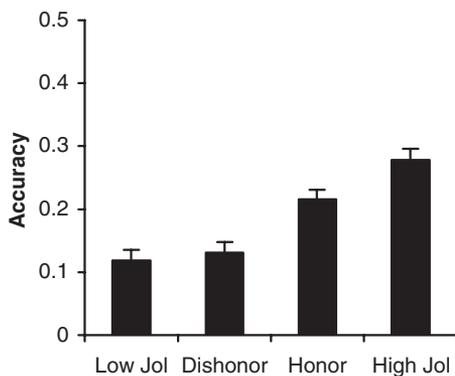


Figure 3. Proportion correct in Experiment 3a. Error bars represent standard errors. Jol = judgment of learning.

honor condition ($M = .21$), $t(35) = 1.01$, $p = .29$. Again, this shows that performance in the honor condition, in which participants controlled their own study, was highly efficient.

Experiment 3b

In all of the experiments presented thus far, the items were presented one by one during the choice phase. In real life, however, study choices are often made from among a set of items presented together, such as when a list of vocabulary words is presented on a page of a textbook. Sequential and simultaneous presentation during choice can lead to different outcomes. Specifically, when people are given easily achievable performance goals, they select easy items with simultaneous presentation, but difficult items with sequential presentation (Thiede & Dunlosky, 1999). One explanation for this is that participants normally choose to study difficult items, but that there is a “shift-to-easier-materials effect,” in which participants change their strategy and study easier items. Simultaneous presentation allows this shift to occur because it allows people to plan their study choices before making them. Item-by-item presentation is less likely to allow the shift to occur because it requires participants to remember which items they have already chosen, creating large working memory demands. Thus, instead of trying to make strategic choices, people revert to choosing difficult items in an effort to try to learn everything. In support of this explanation, the shift-to-easier-materials effect was not apparent in people with low working memory capacity (Dunlosky & Thiede, 2004).

Given that participants in Experiment 3a chose to study easy items with sequential presentation, simultaneous presentation might be expected to heighten this preference. It might also make choice more effective if working memory problems existed in Experiment 3a that could be remedied with simultaneous presentation. Regardless of working memory, participants might also do better simply because simultaneous presentation allows them to examine all of the items before making their choices as well as to go back and change their selections if necessary. This might lead to better performance in the honor condition in part because it avoids the problem of forced choices. Forced choices made it difficult to compare the honor and high-JOL conditions; removing them makes a more direct test of the how well people can do when they control their own study possible. Thus, the goal of Experiment 3b was twofold: to replicate Experiment 3a using simultaneous presentation during choice and to investigate further how well participants can do when they control their own study as compared with a condition (high JOLs) that should lead to excellent performance.

Method

The only methodological difference between Experiments 3a and 3b was in the choice phase. During the choice phase in Experiment 3a, the items were presented one by one. In Experiment 3b, all of the items were presented simultaneously, and the participants selected the ones they wanted to study by clicking on them. They were also allowed to unclick items if they changed their minds. As before, they had to choose exactly half of the items to study.

The participants were 28 Columbia University undergraduates. Four participants had to be replaced because they reported choosing the opposite

of what they wanted to study based on the expectation that their choices would not be honored, leaving 24 participants.

Results

The results of Experiment 3b were very similar to those of Experiment 3a. First, participants made accurate JOLs, based on the gamma correlation between JOL and initial test performance, which was significantly greater than zero ($M = .83$), $t(23) = 37.63$, $p < .0001$.

Participants chose to study the easy items. The mean gamma correlation between JOL and study choice was significantly greater than zero ($M = .45$), $t(23) = 7.10$, $p < .0001$. This was confirmed by self-reports: 20 of the 24 participants reported studying the easy items. The other 4 participants accounted for the three lowest study–JOL gamma correlations, all of which were negative, and the fourth was below the median. These participants did bring the mean down, but because of a negatively skewed distribution, the median gamma correlation (.49) was not substantially higher than the mean (.45). These participants were also hurt by their strategy: They accounted for three of the four lowest honor–dishonor difference scores, and all four were below the median. Their honor–dishonor difference scores were also significantly lower than the scores of the remaining 20 participants, $t(22) = 3.35$, $p < .01$. Thus, most participants chose to study the easy items, but a small minority did not, and they learned less as a result.

The effect of condition was again significant, $F(3, 23) = 4.12$, $p < .01$ ($ES = .15$). Performance showed the same pattern as Experiment 3a and was in the following order: high JOL ($M = .25$), honor ($M = .23$), dishonor ($M = .16$), low JOL ($M = .15$; see Figure 4). The high-JOL condition was better than the low-JOL condition, $t(23) = 3.13$, $p < .01$, and the honor condition was better than the dishonor condition, $t(23) = 1.74$, $p < .05$ (one-tailed). There was no significant difference between the high-JOL and honor conditions, $t(23) = .84$, $p = .41$, or between the dishonor and low-JOL conditions, $t(23) = .23$, $p = .82$. This analysis was repeated including only the 20 participants who selected the easy items. The main effect of condition was significant, $F(3, 19) = 9.56$, $p < .0001$ ($ES = .33$). The order of means was as follows: high JOL (.25), honor (.24), low JOL (.12), dishonor (.11). The high-JOL condition was better than the low-

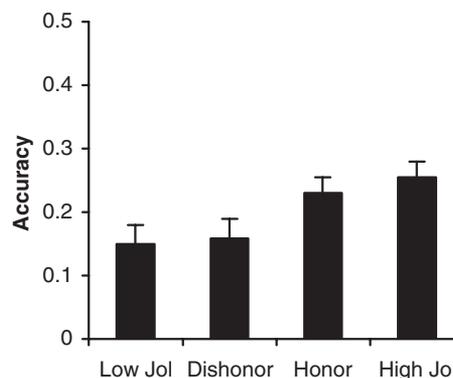


Figure 4. Proportion correct in Experiment 3b. Error bars represent standard errors. Jol = judgment of learning.

JOL condition, $t(19) = 3.68, p < .01$, and the honor condition was better than the dishonor condition, $t(19) = 4.23, p < .001$. The high-JOL and honor conditions, $t(19) = .32, p = .75$, and the dishonor and low-JOL conditions, $t(19) = .09, p = .93$, were not significantly different.

Three questions can be answered based on these data, and the answers all matched those obtained in Experiment 3a. First, the best strategy was to study the easy items because the high-JOL condition was better than the low-JOL condition. Second, participants benefited from controlling their study because the honor condition was better than the dishonor condition. And finally, performance in the honor condition was highly adaptive: The honor condition was not significantly different from the high-JOL condition whether outliers were included or not.

Combined Analyses of Experiments 3a and 3b

To investigate possible differences between sequential and simultaneous presentation, we compared Experiments 3a and 3b. The gamma correlation between JOLs and initial test accuracy did not differ significantly between Experiments 3a and 3b, $t(70) = .10, p = .92$, which is not surprising because both variables in the correlation were measured before the choice phase. There was also no significant effect of experiment on the study choice–JOL gamma correlation, $t(70) = 1.39, p = .17$, although the mean with sequential presentation (Experiment 3a, $M = .32$) was numerically lower than that with simultaneous presentation (Experiment 3b, $M = .45$). The analysis of variance comparing test accuracy across the four conditions also showed no significant main effect of experiment, $F(1, 70) = .42, p = .52$ ($ES = .01$), and no interaction, $F(3, 70) = .84, p = .47$ ($ES = .01$), although the main effect of condition remained, of course, $F(3, 70) = 20.92, p < .0001$ ($ES = .23$). Thus, in this paradigm, sequential versus simultaneous presentation during study choice had no significant effects on metacognition, choice, or performance. The reason for this is probably that, unlike in previous experiments (Dunlosky & Thiede, 2004; Thiede & Dunlosky, 1999), participants did study strategically in the sequential presentation experiment.

Discussion

Experiments 3a and 3b produced four main findings. First, when asked to choose items for study from among a set of items they did not know, people chose to study the easiest items. Second, the strategy that worked best in this experiment was to study the easiest items. Third, not surprisingly given the first two findings, people benefited from controlling their study. Finally, when people were allowed to control how they studied, their performance was highly efficient, and it was not significantly different from performance in the high-JOL condition, which, if RPL is correct, should lead to excellent and possibly even optimal performance. The findings were the essentially the same whether study choices were made one by one (in Experiment 3a) or with all items presented simultaneously (in Experiment 3b).

These findings are consistent with the RPL framework, but inconsistent with DR, in terms of what people actually do as well as what they should do. They show that what people do is the same as what they should: not study the most difficult items, as DR

predicts, but instead study the items just out of reach, that is, the easiest items they have not yet learned, as RPL predicts.

General Discussion

There were three main findings. First, metacognitively guided study led to effective learning. Using a variety of materials and in a variety of situations, people consistently learned more when they were allowed to control their study than when they were not. Second, in terms of how people *should* study, the results supported the RPL model of study-time allocation (Metcalfe, 2002; Metcalfe & Kornell, 2003, 2005). To accomplish the most learning in the least time, people should study the items they do not know and, among them, prioritize the easiest and thus most learnable items. Third, the way people actually *did* choose to study was in line with how they should have, prioritizing easy unlearned items, again supporting RPL. Furthermore, Experiments 3a and 3b showed that, when participants controlled their study, their learning was highly adaptive and not significantly different from learning in the best computer-controlled condition (the high-JOL condition), which should lead to excellent, and possibly optimal, performance according to RPL.

In Experiments 1 and 2a–d, DR and RPL both predicted that people would choose to study the items they did not know instead of the items they knew, and these predictions were supported. In Experiments 3a and 3b, participants were forced to choose among items they did not know, choosing study for some and not for others. In this situation, DR predicts that participants should choose to study the most difficult items. RPL makes the opposite prediction, that participants should prioritize the easiest and most learnable items. The pattern of choices supported the predictions of RPL. Experiments 3a and 3b also included two conditions in which the computer selected items for study based on participants' JOLs, not their study choices; comparing these conditions showed that studying the easiest unlearned items led to better test performance than studying the most difficult unlearned items, again as predicted by RPL.

A new way to conceptualize the choices people face when studying is in terms of learnability: the more learnable an item is at a given time, the more its probability of being recalled on a later test will increase as a result of study. For example, during the study-choice phase of Experiments 3a and 3b, participants had just demonstrated that they did not know the items, so the probability of future recall of any item without further study was approximately zero, but the chance of *increasing* that probability was higher (i.e., learnability was higher) for the relatively easier items than for the more difficult items (as demonstrated empirically by the superiority of the high-JOL condition over the low-JOL condition).

In the past, it has been thought that study decisions are made on the basis of JOLs; for example, both DR and RPL, at least in their simplest versions, make predictions about study decisions as a function of JOLs. The concept of learnability leads to a new way of understanding study decisions. Learnability is conceptually closer to a judgment of *change* in learning (i.e., the amount that learning will increase with practice) than to a JOL itself. Metcalfe and Kornell (2005) have recently hypothesized that people rely on learnability when deciding how long to persist on an item once they have chosen to study it. Their findings suggest that people

continue to study until their jROL is too low, that is, they stop studying when their level of learning stops increasing. Similarly, people may decide which items to study based on judgments of learnability instead of relying directly on JOLs. This is consistent with RPL; under time constraints such as those in the present experiments, RPL predicts that the easiest unlearned items are the most learnable, and that people should, and do, study them.

RPL predictions depend on a person's goals and level of knowledge and the time available for study (Metcalf, 2002; Metcalf & Kornell, 2003). If the goal is to achieve mastery, the most difficult items will have to be given the most attention (and in fact, studying difficult items may result in their moving into the region of proximal learning). Educationally realistic settings in which people have time to study everything to the point of mastery may be the exception rather than the rule, but under those conditions, DR and RPL make similar predictions.

RPL makes two types of predictions—what people actually study and what they should study—and in many situations, the two predictions are essentially the same. There are exceptions, however, that result from errors in monitoring, control, or both. For example, in Experiments 3a and 3b, monitoring resolution was good, and test performance in the self-study condition was close to optimal, but participants demonstrated marked and consistent overconfidence (see Koriat, 1997; Koriat, Sheffer, & Ma'ayan, 2002), overestimating performance across lists by an average of 9.8 percentage points, which can lead to suboptimal study.

Across studies, a small but consistent minority of participants, about 10–20%, demonstrated poor study strategies. These participants' learning suffered as a result. In Experiments 1 and 2a–d, these participants failed to study the items they did not know, a fairly obvious strategy. A slightly more complex strategy, studying the easiest unknown items, was violated at approximately the same rate in Experiments 3a and 3b. Some participants reported a study strategy directly opposite the majority; others reported other strategies, such as choosing randomly. These experiments involved university students, who have enormous amounts of experience studying. Rates of such metacognitive control problems may be higher among younger students.

People with deficient strategies stand to benefit from training in how to study or from learning programs that guide their study. Improving one's approach to studying is perhaps the least effortful way to improve learning because it makes study more efficient without significantly increasing effort. To train people in the relatively simple study strategies described here might be accomplished with little effort, although attempts to do so need to be put into practice and tested empirically. Thus, a practical implication of this research is that training people to study effectively might be enormously helpful, especially for the subset of the people with maladaptive study strategies.

References

- Atkinson, R. C. (1972). Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology*, *96*, 124–129.
- Begg, I. M., Martin, L. A., & Needham, D. R. (1992). Memory monitoring: How useful is self-knowledge about memory? *European Journal of Cognitive Psychology*, *4*, 195–218.
- 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.
- Bisanz, G. L., Vesonder, G. T., & Voss, J. F. (1978). Knowledge of one's own responding and the relation of such knowledge to learning: A developmental study. *Journal of Experimental Child Psychology*, *25*, 116–128.
- Bjork, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical Aspects of Memory: Current Research and Issues: Vol. 1. Memory in Everyday Life* (pp. 396–401). New York: Wiley.
- Calfee, R. C., & Atkinson, R. C. (1965). Paired-associate models and the effects of list length. *Journal of Mathematical Psychology*, *2*, 254–265.
- Cavanaugh, J. C., & Perlmutter, M. (1982). Metamemory: A critical examination. *Child Development*, *53*, 11–28.
- Cull, W. L., & Zechmeister, E. B. (1994). The learning ability paradox in adult metamemory research: Where are the metamemory differences between good and poor learners? *Memory & Cognition*, *22*, 249–257.
- Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition*, *25*, 691–700.
- Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multitrail learning. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *52*, 178–186.
- Dunlosky, J., & Hertzog, C. (1998). 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. 249–275). Mahwah, NJ: Erlbaum.
- Dunlosky, J., Kubat-Silman, A. K., & Hertzog, C. (2003). Training monitoring skills improves older adults' self-paced associative learning. *Psychology and Aging*, *18*, 340–345.
- Dunlosky, J., & Nelson, T. O. (1992). Importance of kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, *20*, 374–380.
- Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect people's self-paced study and yield "labor-and-gain" effects. *Acta Psychologica*, *98*, 37–56.
- Dunlosky, J., & Thiede, K. W. (2004). Causes and constraints of the shift-to-easier-materials effect in the control of study. *Memory & Cognition*, *32*, 779–788.
- Hager, W., & Hasselhorn, M. (1992). Memory monitoring and memory performance: Linked closely or loosely? *Psychological Research*, *54*, 110–113.
- Hasselhorn, M., & Hager, W. (1989). Prediction accuracy and memory performance: Correlational and experimental tests of a metamemory hypothesis. *Psychological Research*, *51*, 147–152.
- Kearney, E. M., & Zechmeister, E. B. (1989). Judgments of item difficulty by good and poor associative learners. *American Journal of Psychology*, *102*, 365–383.
- Kelly, M., Scholnick, E. K., Travers, S. H., & Johnson, J. W. (1976). Relations among memory, memory appraisal, and memory strategies. *Child Development*, *47*, 648–659.
- Kennedy, M. R. T., Carney, E., & Peters, S. M. (2003). Predictions of recall and study strategy decisions after diffuse brain injury. *Brain Injury*, *17*, 1043–1064.
- Kimball, D. R., & Metcalf, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory & Cognition*, *31*, 918–929.
- Koriat, A. (1997). Monitoring ones own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349–370.
- Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased overconfidence with practice. *Journal of Experimental Psychology: General*, *131*, 147–162.
- Lovelace, E. A. (1984). Metamemory: Monitoring future recallability

- during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 756–766.
- Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 663–679.
- Maki, R. H., & Swett, S. (1987). Metamemory for narrative text. *Memory & Cognition*, 15, 72–83.
- Masur, E. F., McIntyre, C. W., & Flavell, J. H. (1973). Developmental changes in apportionment of study time among items in a multitrial free recall task. *Journal of Experimental Child Psychology*, 15, 237–246.
- Mazzoni, G., & Cornoldi, C. (1993). Strategies in study-time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: General*, 122, 47–60.
- Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition*, 18, 196–204.
- Metcalfe, J. (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., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, 52, 463–477.
- Moulin, C. J. A., Perfect, T. J., & Jones, R. W. (2000). The effects of repetition on allocation of study time and judgments of learning in Alzheimer's disease. *Neuropsychologia*, 38, 748–756.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Retrieved August 5, 2003, from <http://www.usf.edu/FreeAssociation/>
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing. *Psychological Bulletin*, 95, 109–133.
- Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning of Swahili-English translation equivalents. *Memory*, 2, 325–335.
- Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, 5, 207–213.
- Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 676–686.
- Nelson, T. O., & Narens, L. (1980). Norms of 300 general-information questions: Accuracy of recall, latency of recall, and feeling-of-knowing ratings. *Journal of Verbal Learning and Verbal Behavior*, 19, 338–368.
- 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.
- Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 76, 1–25.
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255.
- Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty* (2nd ed.). Mahwah, NJ: Erlbaum.
- Son, L. K. (2004). Spacing one's study: Evidence for a metacognitive control strategy. *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.
- Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science*, 3, 315–316.
- Thiede, K. W. (1999). The importance of accurate monitoring and effective self-regulation during multitrial learning. *Psychonomic Bulletin & Review*, 6, 662–667.
- Thiede, K. W., Anderson, M. C. M., & Theriault, 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, and Cognition*, 25, 1024–1037.
- Underwood, B. J. (1966). Individual and group predictions of item difficulty for free learning. *Journal of Experimental Psychology*, 71, 673–679.
- Woodworth, R. S. (1921). *Psychology: A study of mental life*. New York: Holt.
- Yan, W. (1994). Learning ability and memory monitoring. *Intelligence*, 18, 215–229.

Received October 24, 2005

Revision received December 20, 2005

Accepted January 5, 2006 ■

Instructions to Authors

For Instructions to Authors, please consult the January 2006 issue of the volume or visit www.apa.org/journals/xlm and click on the "Instructions to authors" link in the Journal Info box on the right.