

The Dynamics of Learning and Allocation of Study Time to a Region of Proximal Learning

Janet Metcalfe and Nate Kornell
Columbia University

In contrast to the dominant discrepancy reduction model, which favors the most difficult items, people, given free choice, devoted most time to medium-difficulty items and studied the easiest items first. When study time was experimentally manipulated, best performance resulted when most time was given to the medium-difficulty items. Empirically determined information uptake functions revealed steep initial learning for easy items with little subsequent increase. For medium-difficulty items, initial gains were smaller but more sustained, suggesting that the strategy people had used, when given free choice, was largely appropriate. On the basis of the information uptake functions, a negative spacing effect was predicted and observed in the final experiment. Overall, the results favored the region of proximal learning framework.

A number of researchers have argued that the central function of accurate metacognition is to provide people with optimal control of their own learning (Koriat, 2000; Nelson & Dunlosky, 1991; Nelson & Narens, 1990, 1994) by allowing them to allocate their study time effectively. Benjamin, Bjork, and Schwartz (1998) put it this way: “Poor self-monitoring capacity necessarily entails poor selection and execution of relevant control processes: If you do not know what you do not know, you cannot rectify your ignorance” (p. 65). Nelson and Dunlosky (1991) argued, “The accuracy of JOLs [judgments of learning] is critical because if the JOLs are inaccurate, the allocation of subsequent study time will correspondingly be less than optimal” (p. 267). Although these ideas sound reasonable, at the present time, not only are there no data on the relation between people’s metacognitions and their appropriate allocation of study time, there are almost no data about the even more fundamental question of what constitutes optimal allocation of study time—a question of focal concern in this article. There are, however, two theories.

The dominant model of study-time allocation—the discrepancy reduction model—was forwarded by Dunlosky and Hertzog (1998) and is consistent with the views articulated, with slightly different terminology, by a large number of other researchers, including Carver and Scheier (1990), Dunlosky and Thiede (1998), Hyland (1988), Koriat and Goldsmith (1996), Le Ny, Denhiere, and Le Taillanter (1972), Lord and Hanges (1987), Nelson and Narens (1990), Powers (1973), and Thiede and Dunlosky (1999). It proposes that people selectively allocate their time to those items they judge to be the most discrepant from an internal criterion of

being learned, that is, to the items that are judged to be most difficult. This view is also compatible with certain learning theories such as back-propagation (see Rumelhart & McClelland, 1986), in which learning proceeds by first correcting the largest errors and then, on successive epochs, correcting successively smaller errors until a learning criterion is reached.

The discrepancy reduction model does an excellent job of accounting for much of the empirical data on study-time allocation. For example, Son and Metcalfe (2000) reviewed the literature and found 46 separate experimental conditions contained in 19 published reports. Of these, 35 showed that, as the discrepancy reduction model asserts, people allocated their study time to the items either that were objectively most difficult or that they judged to be most difficult (Bisanz, Vesonder, & Voss, 1978; Cull & Zechmeister, 1994; Dunlosky & Connor, 1997; Kellas & Butterfield, 1971; Kobasigawa & Metcalf-Haggert, 1993; Le Ny et al., 1972; Mazzoni & Cornoldi, 1993; Mazzoni, Cornoldi, & Marchitelli, 1990; Mazzoni, Cornoldi, Tomat, & Vecchi, 1997; Nelson, Dunlosky, Graf, & Narens, 1994; Nelson & Leonesio, 1988; Thiede & Dunlosky, 1999; Zacks, 1969; and also see Pelegrina, Bajo, & Justicia, 2000). The predicted monotonic relation between JOLs and study-time allocation was violated in very few participant populations, for example, very young children (Dufresne & Kobasigawa, 1988, 1989; Kobasigawa & Dufresne, 1992; Masur, McIntyre, & Flavell, 1973) and mentally retarded teenagers (Belmont & Butterfield, 1971).

The nature of these exceptions invites the inference that to optimize learning, people should allocate time in accordance with the discrepancy reduction model. The metacognitions in these populations may have been poor, and with these poor metacognitions, the participants may have failed to allocate their study time appropriately (i.e., to the most difficult items). Consequently, their performance may have suffered.

Although this inference has some plausibility, direct empirical evidence for the notion that it is best to selectively study the most difficult items is nonexistent. To our knowledge, no studies have shown that selective allocation of study time to the difficult items results in better performance than does selective allocation to the

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

We thank the James S. McDonnell Foundation and the National Institute of Mental Health (Grant MH60637) for supporting this research. We thank Lisa Son, Hedy Kober, Bridgid Finn, and William Kelemen for their help.

Correspondence concerning this article should be addressed to Janet Metcalfe, Department of Psychology, Columbia University, New York, New York 10027. E-mail: metcalfe@psych.columbia.edu

items that are easy or of medium difficulty. One study by Nelson et al. (1994) suggested that self-choice of items and choice of the “worst-learned” items resulted in best performance. However, the items in this experiment were preselected to be the easiest 36 items of a set of 100 items, so it is not clear whether they should be considered difficult or easy. Furthermore, a study by Nelson and Leonesio (1988) suggests that the allocation of study time to the most difficult items might be counterproductive. They investigated the relation between freely allocated study time, JOLs, and performance, emphasizing either speed or accuracy. Participants spent up to 10 s more per item in the accuracy condition than in the speeded condition. They tended to devote this extra time particularly to the difficult items. But there was very little gain in performance. Unfortunately, insofar as the study was only correlational, it is not known whether the participants could have done better by allocating time to items other than those that they chose. In any event, a large increase in time devoted to difficult items resulted in so little gain that the authors dubbed their results the “labor-in-vain effect” (Nelson & Leonesio, 1988, p. 681).

Furthermore, several recent studies have cast doubt on the generality of the discrepancy reduction model even as a statement of what people do when allocating their study time. In nearly all of the studies that support the discrepancy reduction model, the items were presented one at a time, and participants were allowed to study them for as long as they wanted. There was no ostensible cost to continued study of a difficult item, insofar as it did not mean there would be less time for other items. The exceptions to the rule that people allocate most study time to the most difficult items include experiments in which participants were encouraged to keep their study time brief (Mazzoni & Cornoldi, 1993; Mazzoni et al., 1990) and experiments in which they were free to choose among different items, rather than only having one item presented at a time, so that spending more time on one item meant that less time was left for the others. Thiede and Dunlosky (1999) found that when people were allowed to choose items for study from among all of the items available, they chose the easy items first under conditions of time pressure or when easily attainable goals were set. Son and Metcalfe (2000) found that under free-choice conditions, college students often chose to study the easiest items longer than items judged to be more difficult. And Metcalfe (2002) showed that these effects were systematic. When people were given a very short amount of time, they tended to study the easy and medium-difficulty items preferentially, but as the amount of study time increased, they shifted toward studying more difficult items. When people were more expert in the domain, as compared with being novices, they also showed a shift in their study-time allocation toward the more difficult items. The results of these experiments contradict the predictions of the discrepancy reduction model and point toward a second theory.

The second theory of study-time allocation, the region of proximal learning framework (see Metcalfe, 2002), derives from a number of distinguished theories of human learning, including those of Atkinson (1972a, 1972b), Berlyne (1978), Hebb (1949), Piaget (1952), and Vygotsky (1987). In all of these theories, the concepts and items that are most amenable to learning are those that are just slightly beyond the individual’s current grasp. Items that are already learned and items that are too difficult for the person to master given his or her current cognitive state are considered to be outside the region of proximal learning and

undesirable candidates for study. This theory predicts that when people are allowed to choose among items for study while under time constraints, they should turn initially to those items that are least difficult, so long as they are not yet learned, declining further study of these easiest items only once learning is thought to have occurred. They should then turn to items of slightly greater difficulty, once again spending as much time as is needed (which will, presumably, be longer than the amount of time needed for the easiest items). Only late in the process should they turn to the most difficult items. The dynamic process resulting from the region of proximal learning framework is consistent with the data of Son and Metcalfe (2000), with Metcalfe (2002) described above, and perhaps with those of Thiede and Dunlosky (1999) and Mazzoni et al. (1990). However, while these studies suggest that people might be trying to hone in on what they believe is their own region of proximal learning, they provide no evidence that what they chose to study was optimal.

Indeed, we were able to find only one study that systematically investigated whether people devote their study time to the materials of appropriate difficulty: an early study by Atkinson (1972a, 1972b). Basing his experiments on a Markov model of human learning by Calfee and Atkinson (1965) in which to-be-learned materials could be in one of three stages—permanently learned, transitory, or unlearned—he posited that the most effective strategy is to devote time to items that are in the transitory state. According to this model, the learning objective is to transfer as many items as possible into the permanently learned state for later delayed recall. To do this, given that all items must go through the transitory state, one should selectively study those items that are already in the transitory state.

Using a computer algorithm, Atkinson (1972a, 1972b) was able to ascertain which of the items—English–German vocabulary pairs—were in each of the three states. He then had the computer assign study opportunities on the basis of the state of learning of the items or allowed people to select items themselves. When participants were allowed to determine what they studied, they did 53% better than when the computer chose items for study randomly. The performance gain with self-selection was nearly identical to that exhibited when the computer randomly selected among any items that were not yet permanently learned (i.e., from both transitory and unlearned items). The most impressive finding was that when the computer preferentially selected the transitory items for people to study, participants gained 108% over the random control. Atkinson (1972a) concluded that people are far from optimal in their self-selected study-time allocation and that their metacognition and control processes were faulty:

One way to avoid the challenge and responsibility of developing a theory of instruction is to adopt the view that the learner is the best judge of what to study, when to study, and how to study. I am alarmed by the number of individuals who advocate this position despite a great deal of negative evidence. . . . My data, and the data of others, indicate that the learner is not a particularly effective decision maker. (p. 930)

Without disputing Atkinson’s conclusions about people’s decision-making and metacognitive capabilities in general (see Metcalfe, 1998), one concern with his conclusion about their allocation of study time is that the instructions given to his participants in the self-determined-study condition may have unfav-

vorably biased their strategy. Participants “were told that their trial-to-trial selection of items should be done with the aim of mastering the total set of vocabulary items” and “that it was best to test and study on words they did not know rather than on ones already mastered” (Atkinson, 1972b, p. 125). It seems plausible that participants may have interpreted these instructions as telling them to study equally all items that were as yet unlearned (regardless of difficulty) but to ignore the items that were already learned. The fact that the self-determined-study condition and the equal-parameter condition (in which transitory and unlearned items were studied the same amount) produced nearly identical data on every trial suggests that people may not have felt completely free to study whatever they liked but rather that they mimicked the equal-parameter condition. Atkinson’s study notwithstanding, then, we do not know whether people’s study-time allocation is appropriate.

In the first two experiments, we examined the questions of to which items people devote their time, in what order, and whether that appropriation is reasonable. In Experiment 1, as well as detailing people’s patterns of study-time allocation and of item switching, we compared the amount of time spent studying with a measure of the rate of learning return per second of study. The idea behind this latter analysis was that the most study time should be allocated to items that returned the most learning per second of study. We used a paradigm (Metcalf, 2002) in which people were given three study possibilities—one easy pair, one medium pair, and one difficult pair—at a time and had to click the computer mouse on a cue word of the pair to see the target. The total amount of time for each triad was systematically varied. In the second experiment, rather than leaving the choice up to the participant, we manipulated the amount of time allowed to easy, medium, and difficult pairs in an effort to determine which emphasis was best. In later experiments, we investigated certain implications of the findings of the first two experiments.

Experiment 1

In the first experiment, we administered a pretest and then allowed people a self-determined learning opportunity. Participants were then tested on their learning and, after we excluded words that were known on the pretest, the percentage of gain per second of study in each treatment combination was computed and compared with self-determined allocation of study time. The hypothesis was that people studying optimally should devote the bulk of their study time to the items that gave the greatest rate of return per second of study, and the return rates should mirror people’s study-time allocation patterns.

Method

Participants. Participants were 12 Columbia University students who received partial course credit for their participation. In all of the experiments in this series, each participant was tested individually on an iMac computer. Participants were treated in accordance with the ethical standards of the American Psychological Association. Participants were not excluded from participating because of knowledge of Spanish, but they were given a pretest so their fluency with the language could be assessed. Although in this article we do not report details concerning participant

Spanish language expertise in any experiments in this series, they are available from Janet Metcalfe.

Design. The experiment was a 3 (difficulty of the materials: easy, medium, difficult) \times 3 (time allowed per triad: 5 s, 15 s, or unlimited time up to 60 s) within-participants design with 16 word pairs per treatment combination. The order of the timing conditions, which was a blocked variable, was counterbalanced across participants.

Materials. The materials were 144 English–Spanish vocabulary pairs from Metcalfe (2002) that were divided into three sets of 48, one for each of the three difficulty conditions. The 48 easy pairs tended to be cognates; the medium pairs could be mediated by recourse to other known words in English; the difficult pairs were constructed such that we were unable to come up with any mediator.

Procedure. Participants were informed that they would be asked to learn 144 English–Spanish vocabulary pairs but that before doing so, they would be given a pretest. On the pretest, which was given on paper, all 144 English words were listed, and the participants were asked to write in the Spanish translation of any word they knew. They were then instructed that during study, on each trial, three English words would be presented on the computer screen with a question mark button below each, and that the Spanish translation of a given word would be displayed when they clicked on the question mark button below the word. The Spanish translation remained on the screen until they clicked on another question mark button or the study time for that triad of pairs elapsed. Participants were informed that the items on the left were relatively easy, those on the right relatively difficult, and those in the middle of intermediate difficulty. They were told that they were free to allocate their study time however they wished, including visiting the same item multiple times. They were also informed of the study time allowed for each triad in each block of the study phase. During the 5-s and 15-s blocks, the computer automatically proceeded to the next trial when the time period elapsed, whereas in the unlimited-time block, a *next triad* button was presented along with each triad. In this block, the computer went on to the next triad automatically if the participant did not click the *next triad* button for 60 s.

When the participant had completed the study phase, the computer administered a test—an English word was displayed, and the participant was asked to type the Spanish translation and hit the *return* key. The order of presentation during the test was random. Participants were told that they could change their answers up until they hit the *return* key and that their data would be scored both strictly and leniently, so if they had some idea of what the words might be, they should type in their answers, even if they were not entirely certain. They were similarly encouraged to guess on the pretest.

Results

In the analysis of variance (ANOVA) data presented throughout this article, a cutoff value of $p < .05$ was chosen as the criterion for significance. Tukey tests were used as post hoc tests throughout on simple effects. Post hoc t tests were computed on individual comparisons on interaction terms, and the p values for these are reported directly, because there is no generally agreed-on authoritative reference for what the alpha level should be. Partial eta squared was used as the measure of effect size and is designated *ES* in this article. The recall data were computer scored using an algorithm, created by Brady Butterfield, which assigns a value between .00 and 1.00 based on the extent of letter overlap between the response and the correct answer. Generally, items with scores over .75 tended to be what a human scorer would call spelling mistakes and were therefore reassigned a value of 1, whereas all

others were assigned a value of 0. These scores were then averaged over the 16 trials.^{1,2}

Recall. In the analyses that follow, items answered correctly on the pretest were excluded—an average of 6.00 easy, 1.58 medium, and 0.08 difficult items per participant. Recall means are shown in Table 1.

The Time Allowed \times Difficulty interaction was significant, $F(4, 44) = 2.91, MSE = 0.01, ES = .21$. This interaction qualified main effects of both time allowed, $F(2, 22) = 10.50, MSE = 0.02, ES = .49$, and difficulty, $F(2, 22) = 123.44, MSE = 0.03, ES = .92$. These results replicate those in Metcalfe (2002) showing that easier items were easier, more time resulted in better performance, and more time was ineffective in improving performance on the most difficult items, although it did help the easy and medium items.

Post hoc Tukey tests confirmed that participants did better on easy pairs than either medium or difficult pairs. The difference between medium and difficult pairs was also significant. Post hoc tests also confirmed that performance was worse in the 5-s condition than in either the 15-s or the unlimited-time condition, although the difference between the 15-s and unlimited-time conditions was not significant. To further investigate the interaction, we compared recall performance in the 5-s and unlimited-time conditions for each level of difficulty. For easy pairs, the difference, favoring the unlimited-time condition, was significant, $M = 0.21, t(11) = 2.69, p = .021$. For medium pairs, this difference was also significant, $M = 0.16, t(11) = 3.95, p = .002$. For difficult items, however, the difference was not significant, $M = 0.04, t(11) = 1.74, p = .111$: More time did not result in better performance. The null effect of increased study time on memory for the difficult items corroborates Nelson and Leonesio's (1988) finding of a labor-in-vain effect. Our data also show, though, that labor was not in vain on the easy and medium items.

Study-time allocation. Figure 1 gives the proportion of time allocated to pairs that were easy, medium, and difficult as a function of time allowed. There was a Time Allowed \times Difficulty interaction, $F(4, 44) = 3.40, MSE = 0.01, ES = .24$, as well as a main effect of difficulty, $F(2, 22) = 6.39, MSE = 0.03, ES = .37$. The main effect showed that people tended to devote more time to the pairs of medium difficulty than to either the easy or the difficult pairs. The interaction replicates Metcalfe (2002), showing that as more time was allowed, people shifted to more difficult pairs. Post hoc tests on proportion of time allocated revealed that more time was spent on the medium pairs than on the easy pairs, and more time was spent on the medium pairs than on the difficult pairs. The difference in time spent between the easy pairs and the difficult pairs was not significant.

Dynamic patterns of selection of items. We also analyzed the order in which easy, medium, and difficult items were studied by

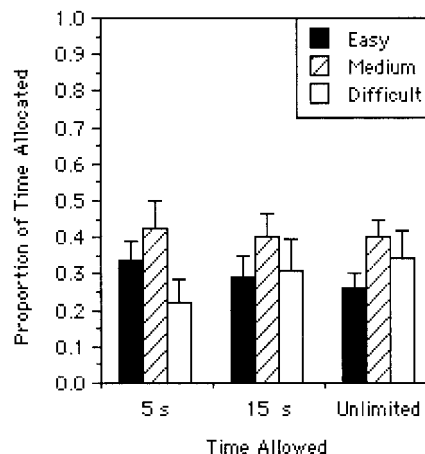


Figure 1. The proportion of time devoted to easy, medium, and difficult items in Experiment 1. Error bars indicate 95% confidence intervals, which are based on individual cell means here and throughout.

calculating the percentage of trials on which each type of item was selected first, second, and third. Participants only occasionally made more than three selections per trial. Data from the three time conditions in the experiment were collapsed. As can be seen from Figure 2, easy items tended to be chosen first. This result is consistent with those of Thiede and Dunlosky (1999). Items of medium difficulty were chosen second. The difficult items tended to be chosen third, although there was also some tendency for people to return to the easy items on their third choice. (Note that the proportion data do not add up to 1 because of the cases—not explicitly enumerated—in which the person did not get to the choice in question within the time limits of the trial.)

Rate of return for time spent. This analysis was directed at determining how much learning resulted from study time devoted to the different difficulty levels. We computed a return rate measure as follows: For each treatment combination (and separately for each participant), first the items that were correct on the pretest were excluded. Then the proportion correct was computed, and this was divided by the average amount of time spent on items in that condition, yielding the proportion recalled per second spent studying. As can be seen from Figure 3, there was an effect of difficulty, $F(2, 22) = 80.45, MSE = 0.01, ES = .88$, such that, as post hoc analyses showed, time spent on the easy items resulted in significantly more gain than did time spent on the medium or difficult items. There was also a main effect of time allowed, $F(2, 22) = 13.50, MSE = 0.01, ES = .55$, showing that there was more

Table 1
Mean Proportion of Items Recalled From Experiment 1

Time	Item type		
	Easy	Medium	Difficult
5 s	.50	.07	.01
15 s	.64	.22	.02
Unlimited	.71	.23	.04

¹ Both pretest and test performances were also measured by a more strict criterion. The patterns of results paralleled those reported here (with a few exceptions in which one measure but not the other fell short of the .05 significance criterion). Details of the strict data are available from Janet Metcalfe.

² Masson and Loftus (in press) have contributed a method of computing confidence intervals that removes between-subject variability in designs such as ours. Unfortunately, we were unable to use their methods because of violations of their assumption of similar error terms for main effects and interactions. We were thus forced to use the more conservative standard confidence intervals.

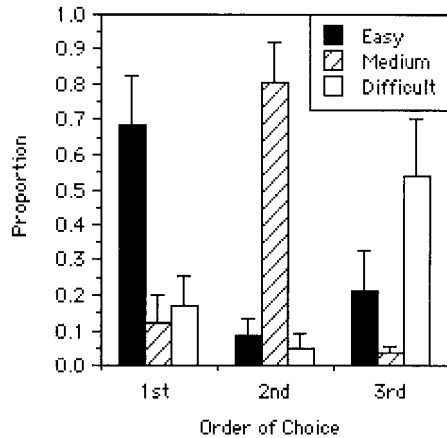


Figure 2. The order of choice of items in Experiment 1. Error bars indicate 95% confidence intervals.

gain per second when less total time, rather than more total time, was allowed. (This effect did not reach significance with stricter scoring, $F[2, 22] = 2.22, p = .13$.) Post hoc analyses showed that there was more gain per second with 5 s of study time than with either 15 s or unlimited time and that the gains with 15 s and unlimited time were not significantly different.

Finally, there was a significant Time Allowed \times Difficulty interaction, $F(4, 44) = 12.08, MSE = 0.01, ES = .52$. When a short amount of time was allowed, time spent on the easy items was especially effective. When more time was allowed, time spent on the easy items was less differentially effective as compared with time spent on the items of medium difficulty. To understand the locus of this interaction in more detail, we first computed the difference between the easy and medium conditions in terms of gain per second. This difference, when the 5-s and 15-s conditions were contrasted, was significant, $t(11) = 4.18, p = .002$, showing that the advantage in gain per second of the easy over the medium condition was larger in the 5-s condition than in the 15-s condition. There was no difference when the 15-s and the unlimited-time condition were compared, however. In summary, time spent on the easy items was especially effective when there was little total time. With longer total time, the advantage to spending it on the easy items diminished.

Correlation between time allocated and recall. We computed a point-biserial correlation between time spent studying individual items and the probability of recall of those items for each participant in each of the time conditions for the easy items. (Unfortunately, the medium and difficult item conditions had to be eliminated because of too few observations.) It seemed self-evident that these correlations should be positive, that is, more time spent studying should lead to a higher probability of recall. Surprisingly, none of the correlations was significantly different from 0, and there was no difference among the three study-time conditions (all $F_s < 1$). More time spent studying the easy items did not increase the probability of learning.

Discussion

These data are inconsistent with the discrepancy reduction model, both in terms of what people should study and what they

did study. They did not selectively devote their study time to the most difficult items but rather to items that were of medium difficulty, replicating earlier studies (Metcalf, 2002; Son & Metcalfe, 2000). Furthermore, there is no indication in these data that selective allocation of time to the difficult items helped performance overall or even on those items themselves. First, when the amount of time was manipulated and people did, therefore, give more time to the most difficult items (on average, 0.71 s in the 5-s condition, as compared with 3.56 s in the 15-s condition and 8.61 s in the unlimited-time condition—a more than 10-fold increase), performance on those difficult items themselves registered no improvement. The additional time did substantially help performance on the easy and medium items, in contrast. Second, the measure of return per second of study indicated no returns (less than 1% per second) for the difficult items. Selectively devoting time to the difficult pairs, rather than being an adaptive strategy, was, indeed, labor in vain. The discrepancy reduction model is undermined by these data.

But perhaps even the devotion of time to the medium items—which were the ones on which people spent most time—may not have been well spent. The return-per-second measure, taken alone, suggests that people should not have studied the medium items for so long and should, instead, have given even more time than they did to the easy items.

Although this might be the case, and we will investigate this possibility further in the next experiment, there are some reasons for doubt. First, it bears pointing out that because this first study was essentially correlational, we do not know, in principle, whether people would have benefited from further study of the easy items or whether more time spent on them might have been wasted. Additionally, two factors in the data from Experiment 1 suggest that people might not have reaped the kinds of additional gains suggested by the return-per-second measure had they devoted further study time to the easy items. First, the zero correlations between study time and performance on the easy items suggest that giving additional time on these items might not necessarily improve performance. Although we acknowledge that there may be factors that could obviate a straightforward interpretation of this lack of correlation within time conditions, it was

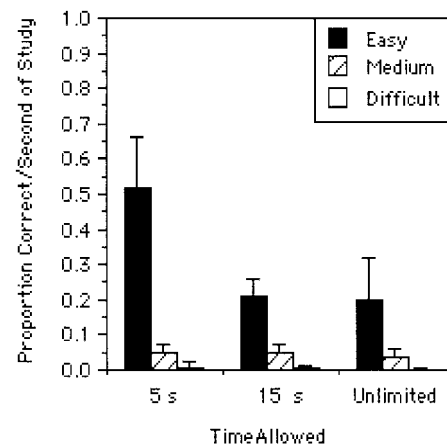


Figure 3. Proportion correct per second of study for the easy, medium, and difficult conditions. Error bars indicate 95% confidence intervals.

nevertheless surprising that an increase in time spent on the easy items did not result in increased performance. Second, on the return-per-second analysis, the interaction between difficulty and time allowed weighs in against the idea that further time spent on the easy items would have resulted in larger performance gains. When a very short amount of total time was allowed, the benefit to spending time on the easy items was large, but as the total time allowed increased, the benefit to spending it on the easy items decreased while the benefit to spending it on the medium-difficulty items stayed about the same.

A trenchant observation by Kahneman (1973; also see Koriat, Sheffer, & Ma'ayan, 2002) also suggests that, in contrast to the return-per-second data, people's strategies of selectively studying the items of medium difficulty and not devoting too much time to the easy items may not have been misguided. Kahneman argued that certain materials only allow so much attention. He said, "the subject simply cannot try as hard in a relatively easy task as he does when the task becomes more demanding" (Kahneman, 1973, p. 14). Perhaps the easy items were of this variety, and their capacity to be studied was quickly exhausted. For these reasons, we decided to investigate experimentally, rather than just correlationally, whether devoting time to easy, medium, or difficult items would result in the best performance.

Experiment 2

We tested three competing hypotheses. The first was that the return-per-second data from Experiment 1 indicated what the person should study. If this was so, then people should have given much more time than they did to the easy items. The second hypothesis was that people basically behaved appropriately in Experiment 1. Perhaps there was some reason why devoting more time to the easy items would not have resulted in more learning and, therefore, people were correct to turn to the medium-difficulty items. The third hypothesis was that the large study-time allocation literature indicating that people frequently devote their time to the most difficult items provides a sketch of the optimal strategy. We doubted that this third possibility could be correct, given the results of Experiment 1, but in light of the fact that 35 experiments using a paradigm different from the one in Experiment 1 had shown that this is what people do, it seemed prudent to include the third hypothesis as a possibility.

The basic idea of the experiment was that item pairs that were easy, medium, or difficult would be shown to participants, but rather than having them control how much time they would spend studying each, we would control the time spent—assigning the majority of the time to either the easy items, the medium items, or the difficult items. Pilot studies showed that we could not adequately control the timing of study while maintaining the spatial layout of Experiment 1, so we presented the pairs one at a time for the amount of time designated by the experimental condition but in the same position on the screen (i.e., in the center).

Method

Participants were 12 Columbia University students who were not fluent Spanish speakers and who received partial course credit for their participation. The materials were the English-Spanish pairs used in Experiment 1.

Participants were given a pretest on all of the words so that already known translations could be eliminated from the analyses. They were encouraged to guess during both the pretest and the final test if they did not know the answer to a question.

All pairs were presented in the same location so that the presentation of the new cue-target pair masked the preceding pair. They were always presented in the order easy, then medium, then difficult. There were three conditions; in each condition, either the easy, the medium, or the difficult items were emphasized. We allowed the two unemphasized conditions 1 s each. The emphasized condition was given 10 s. Thus, for example, the timing in the emphasize-easy condition was 10 s per easy pair, 1 s per medium-difficulty pair, and 1 s per difficult pair. The design, then, was a 3×3 within-participants factorial design where the variables were difficulty (easy, medium, or difficult) and emphasis (easy, medium, or difficult). After the presentation of the words during study, participants were tested on all of the words, one at a time, in random order. After each response, they were asked for a confidence rating.

Results

Items answered correctly on the pretest were excluded from the analyses. An average of 14.08 easy, 2.17 medium, and 0.08 difficult items were excluded per participant.

There was a significant effect of emphasis, $F(2, 22) = 4.39$, $MSE = 0.01$, $ES = .29$ (see Figure 4). The emphasize-medium condition produced the best results, $M = 0.32$. The mean for the emphasize-easy condition was 0.27, and the mean for the emphasize-difficult condition was 0.25. Tukey tests showed that the difference between the emphasize-medium condition and the emphasize-difficult condition was significant, whereas the other differences did not reach the criterion for significance.

The effect of difficulty was significant, $F(2, 22) = 160.15$, $MSE = 0.02$, $ES = .94$. Tukey tests showed that participants recalled the easy items better than they did both the medium and the difficult items, and they recalled the medium items better than they did the difficult items. The Emphasis \times Difficulty interaction

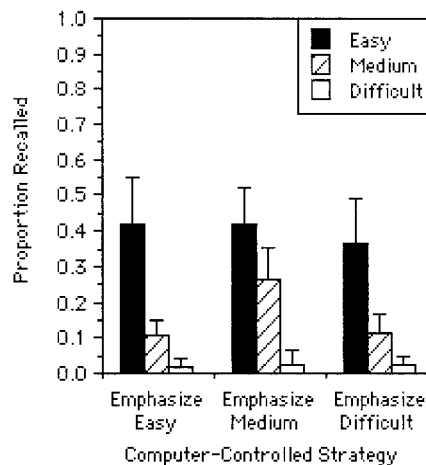


Figure 4. Proportion of easy, medium, and difficult items recalled when the easy items were allocated most study time (10 s, 1 s, 1 s, respectively), when the medium items were allocated most study time (1 s, 10 s, 1 s, respectively), and when the difficult items were allocated most study time (1 s, 1 s, 10 s, respectively), from Experiment 2. Error bars indicate 95% confidence intervals.

was significant, $F(4, 44) = 3.31$, $MSE = 0.02$, $ES = .23$ (although not with stricter scoring, $F[4, 44] = 1.90$, $p = .13$). Post hoc analyses showed that medium items were remembered better when they were emphasized than when they were not, $t(11) = 3.89$, $p = .003$. This difference was not significant for easy items, $t(11) = 1.03$, $p = .327$, or for difficult items, $t(11) = 1.46$, $p = .172$. Thus, only the medium items benefited from additional study time, and the benefit to them showed up as an overall improvement in performance.

Discussion

In this experiment, the best strategy was to selectively allocate time to the items of medium difficulty. The results of this experiment offered no comfort to the discrepancy reduction model: The selective allocation of large amounts of time to the most difficult items was not beneficial.

Although this study did not show an advantage to allowing more than 1 s to the easy items either overall or, indeed, when measured by performance on the easy items themselves, a follow-up experiment using a similar design but in which the times went from 0.5 s to 10 s (rather than from 1 s to 10 s) showed some advantage (which was marginally significant, $F[2, 34] = 2.52$, $MSE = 0.01$, $ES = .13$, $p = .095$, with the scoring criterion reported here, but which was significant with a stricter scoring criterion, $F[2, 34] = 6.45$, $MSE = 0.01$, $ES = .27$) to giving more time than 0.5 s to the easy items. These results, then, conjoined with the results of Experiments 1 and 2, suggest that whether or not additional study time will enhance memory is a joint function of exactly how much study time has already been allocated and the nature of the items. If the items are easy, some small amount of additional study time over a very short baseline may be helpful. If the person has already spent a second or so on an easy item, however, additional time may be wasted. If the items are of medium difficulty, memory may continue to benefit from increases in study time over a larger range. If the items are truly difficult, though, even large increases in study time may have little effect. Delineating the information uptake curves (i.e., the gain in terms of memory percentage increase over the interval in question divided by the amount of time in the interval) that show the relation between memory performance for various item types and the amount of time spent studying is necessary if one is to understand how much time should be devoted to varying item types, and when the individual should decline further study as unproductive.

Experiment 3

The third experiment was designed to investigate the information uptake functions parametrically. We were especially interested in determining whether these functions could resolve the question of why, when the return functions in Experiment 1 had shown that the greatest return per second of study resulted from the easiest items, the results of Experiment 2 indicated that it did not behoove people to devote much more time to them. The conjecture that we explored was that information uptake of the easy items plateaus quickly; that is, there is a considerable benefit from a short amount of study of those items, but additional study on them does not improve performance proportionally. In contrast, infor-

mation uptake of medium-difficulty items was hypothesized to be less steep initially but more sustained over time.

We varied the study time on items of easy and medium difficulty using a new set of English–Spanish materials. We did not include the most difficult items because they were showing no effects whatever of differences in study time in the earlier studies. The time needed to learn these items appeared to be so long as to be intractable in our experimental setting. Items were presented for varying amounts of time and then participants' memory for those items was tested. The two experiments presented in this section were essentially the same (and hence provide a replication of the basic results) except that in Experiment 3A participants were not given a pretest, whereas in Experiment 3B they were.

Method

Participants, design, and materials. Participants were 33 and 29 Columbia University students in Experiments 3A and 3B, respectively, who received partial course credit for their participation.

The basic experiment was a 2 (difficulty of the materials: easy or medium) \times 6 (time allowed per pair: 0, 0.5, 1, 2, 4, or 8 s) within-participants design with 16 English–Spanish pairs per treatment combination. The order of the difficulty conditions, which was a blocked variable, was counterbalanced across participants. In Experiment 3B, the pretest allowed already known items to be eliminated.

The materials were 192 English–Spanish vocabulary pairs, none of which were used in the previous experiments. The new list of words was created because we needed a larger list for this experiment and because it allowed us to avoid screening for and eliminating participants who had run in the previous Spanish–English experiments. The words were chosen to be similar in difficulty to the corresponding words from the previous experiments.

Procedure. Word pairs were presented in blocks of 48 words each. During each block, the words were presented for study, there was a 30-s distractor task in which participants counted backward by 3s from a three-digit number, the words were tested, and then the next block was presented. There were four blocks—two made up of easy words, the other two of medium-difficulty words—and the order of the blocks was counterbalanced.

We were concerned that during study, if we allowed brief presentations to be followed by long ones, participants would continue to think about the briefly presented item while the longer ones were being presented (hence offsetting our time manipulation). The uneven rhythm of random presentation also made study difficult and strange. We did not want to present each of the durations in a separate block because we did not do so in the preceding experiments and because we wanted to equate mean time between presentation and test across conditions. Therefore, during study, the words were presented in scalloped cycles of the following order: 8 s, 4 s, 2 s, 1 s, 0.5 s, 0.5 s, 1 s, 2 s, 4 s, 8 s. This cycle then repeated until all 40 words had been shown. (The 0-s condition is not listed because it was not presented during study.) Before the experiment began, participants studied a practice block of Spanish–English pairs, which were not tested, to familiarize them with this presentation timing.

The order of the words during the test was random. Participants were asked to type in the Spanish translation of the English word that was presented and were told that if they were not sure of the answer, they should guess.

Results

The similarities in the results of the two experiments were striking. They were analyzed separately because the items that were correct in the pretest—an average of 21.48 easy and 6.93

medium-difficulty items per participant—were eliminated in Experiment 3B, and there was no pretest in Experiment 3A.

In Experiment 3A, there was a main effect of difficulty, $F(1, 32) = 474.59$, $MSE = 0.05$, $ES = .94$, showing better performance for the easy items. The easy items were recalled about 50 percentage points higher than medium-difficulty items, overall. There was also an effect of study time, $F(5, 160) = 70.93$, $MSE = 0.01$, $ES = .69$, showing that performance improved with time. The most interesting result, from our perspective, was the Difficulty \times Study Time interaction, $F(5, 160) = 3.51$, $MSE = 0.01$, $ES = .10$. In Experiment 3B, the main effects of difficulty, $F(1, 27) = 674.42$, $MSE = 0.03$, $ES = .96$, and of time allowed, $F(5, 135) = 74.60$, $MSE = 0.01$, $ES = .73$, were significant, as was the crucial Difficulty \times Time Allowed interaction, $F(5, 135) = 5.87$, $MSE = 0.02$, $ES = .18$.

To investigate this central interaction—the information uptake functions themselves—in more detail (and to eliminate differences at the y-intercept, presumably attributable to guessing), we computed a measure of performance gain per second of study in each of the five intervals: (1) from 0 s of study to 0.5 s, (2) from 0.5 s to 1 s, (3) from 1 s to 2 s, (4) from 2 s to 4 s, and (5) from 4 s to 8 s. For each interval, we computed the difference in recall performance from the second study time to the first. We then divided performance gain by the number of seconds in that interval to arrive at performance gain per second. Figure 5 shows these results for both Experiments 3A and 3B. In Experiment 3A, the rate of information uptake for the easy items far exceeded that of the medium-difficulty items in the first interval, that is, between 0 and 0.5 s, $t(32) = 2.86$, $p = .004$ (one-tailed). However, at all subsequent intervals, it favored the medium-difficulty items. For the second interval, the difference favoring the medium-difficulty items was significant, $t(32) = -1.82$, $p = .039$ (one-tailed), as was also the case for the final interval, $t(32) = -1.75$, $p = .045$ (one-tailed).

In Experiment 3B, the advantage for the easy items over the medium-difficulty items in the first interval was slightly smaller than it had been in Experiment 3A and was not quite significant, $t(28) = 1.48$, $p = .075$ (one-tailed). (It was significant with stricter scoring, however: $t[28] = 2.23$, $p = .017$ [one-tailed].) Again, rate of uptake favored the medium-difficulty items at the longer intervals, and significantly so at the fourth interval (i.e., between 2 s and 4 s), $t(27) = -2.52$, $p = .009$ (one-tailed).³ The advantage of the medium-difficulty items over the easy items was marginal in the fifth interval, $t(27) = -1.65$, $p = .055$ (one-tailed).

Discussion

The information uptake curve for the easy items indicated that there was an initial period of fast learning but that the amount of learning diminished very rapidly as more time was spent. This finding is, of course, related to Ebbinghaus's (1885/1964) observation on memory with repetitions—that learning is negatively accelerated (and see Anderson & Schooler, 1991, for an updated discussion; see also, e.g., Stevens, 1962, for possibly related psychophysical data). The uptake curve for the medium-difficulty items neither was as high initially as the easy-items curve, nor did it diminish as much.

These uptake functions make sense of some of the puzzling findings from Experiment 1: that the benefit per second of study of

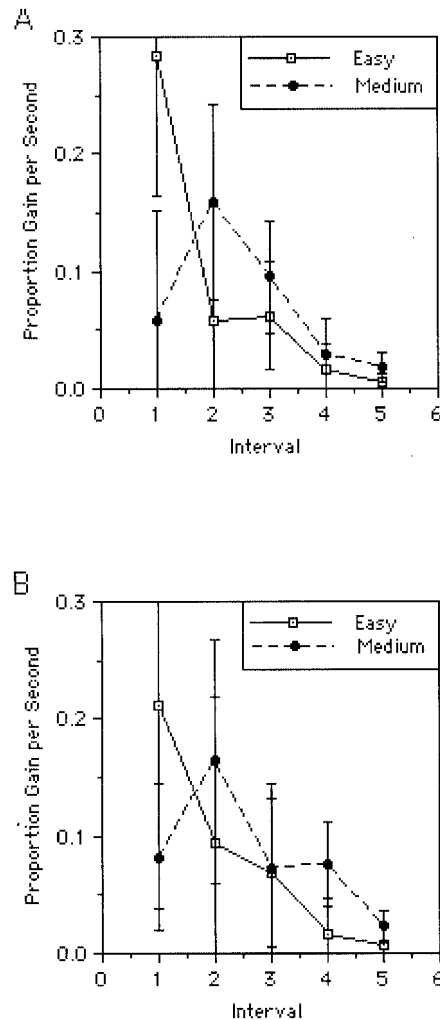


Figure 5. Information uptake functions, that is, gain in recall per second of study, as a function of study interval for the easy and the medium-difficulty items in Experiment 3A (A) and Experiment 3B (B). Error bars indicate 95% confidence intervals.

easy items was greater when there was a short amount of total study time but not as marked when there was more study time. They also make sense of the lack of correlation between recall performance and the amount of study time allocated to easy items. It seems intuitive that the results of Experiment 2—that it was better to devote most time to the medium-difficulty items—also follow, particularly given that about 1 s of study on the easy items was enough to allow people to reach a learning plateau. More time on those easy items would be wasted, according to the uptake curves, whereas time on items of medium difficulty would be better spent. In addition, these curves are consistent with the finding in the follow-up study to Experiment 2 that additional study time over a very short baseline of 0.5 s did benefit the easy

³ In the fourth and fifth intervals, data from 1 participant could not be used because he had a perfect score on the easy items on the pretest, hence the 27 degrees of freedom.

items. And the information uptake functions also suggest that the strategy that people had adopted in the first experiment—of choosing to study the easy items first but then not spending too much time on them before turning to the medium-difficulty items—was well founded. These curves suggest that the individual should initially study the easy items preferentially but only for a very short amount of time. Then there should be a preference for medium-difficulty items, as the rate of gain per second of study for medium items (especially when they have not been studied at all) surpasses that for easy items. The findings of the first two experiments, then, fall into place, given the uptake functions observed in this experiment.

Experiment 4

As the foregoing experiments suggest, the question of appropriate study-time allocation is a dynamic one concerned not only with what people should study but also with how long they should persevere before shifting to something else. The issue that is addressed in this final experiment is: When is it advantageous to break off study of one item in favor of another? The decision about when to discontinue study of particular items is one that the participants in Experiment 1 faced directly, insofar as their task can be construed as determining not how long, in total, to spend on a particular item but rather when information uptake on a particular item was no longer paying off and thus when they should shift their focus to something else. As we have seen, they chose to study easy items first but then to quickly shift to the items of medium difficulty, on which they persevered for longer.

In general, once information uptake has plateaued, there should be diminishing returns for continued study of that item, at least at that time. In contrast, if the learning of an item has not come close to a plateau, then the person, presumably, should persist in studying that item at that time. These conjectures, based on the information uptake functions of Experiment 3 about when one should persist and when one should decline further study, have straightforward implications for the spacing effect.

Massed practice can be viewed as continued study on the same item at the time of its initial presentation—allowing more time for information uptake. Spaced practice can be viewed as declining to continue to study immediately (in favor of returning to an item later when some forgetting has taken place). We suggest that rather than being a ubiquitous benefit to learning, as has often been assumed, spaced practice should be advantageous if the learning on the first study trial has plateaued. But if it has not—which should occur with moderately difficult materials at short presentation rates—then massed practice should be better, because it allows the person to continue studying while the study is still effective and does not prematurely interrupt the learning process.

A large literature indicates that when items are repeated during study, they are remembered better when their presentations are spaced apart than when they are massed (Greene, 1989; Johnston & Uhl, 1976; Madigan, 1969; Melton, 1967; Whitten & Bjork, 1977; and see especially Bahrick, Bahrick, Bahrick, & Bahrick, 1993, who studied spacing in second language vocabulary learning, and Bjork & Allen, 1970, for theoretical discussion). One exception to this otherwise general rule was provided in an experiment by Waugh (1970), in which she found a spacing effect at a 4-s rate of presentation but no effect at a 1-s rate, a result providing

encouragement to our hypothesis. Landauer and Bjork (1978) demonstrated that under conditions where there is no feedback (that is, the target item is not presented at all during the second and subsequent repetitions of the cue), expanding rehearsal practice—in which the spacing of subsequent repetitions is gradually increased—results in better performance than does evenly spaced practice. But the advantage they demonstrated with this method appears to be due to the fact that participants receive no benefit of repetition, spaced or otherwise, on the items they cannot retrieve, and there are more such items with wide spacing. Pashler (2002) has shown that when feedback is given in a similar situation, spaced practice is better than is expanding rehearsal. (The only other “exception” to the spacing effect has been observed when the test occurs very soon after the last presentation of massed pairs; Glenberg, 1976, 1979. But it is generally accepted that this apparent reversal of the spacing effect is not genuine but rather due to the confound of lag to test.) In the remainder of the literature, the spacing effect has been found to occur in a wide variety of learning and remembering tasks and with a wide variety of materials (Glenberg, 1979; Hintzman, 1976), and it has been studied extensively (for reviews, see Crowder, 1976; Greene 1989; Hintzman, 1974; Russo, Mammarella, & Avons, 2002). The spacing effect has been found to occur not only with humans but also with creatures as lowly as the honeybee (Menzel, Manz, Menzel, & Greggers, 2001) and *Aplysia* (Carew, Pinsker, & Kandel, 1972). As such, it is thought to be a fundamental principle of learning. Despite the seeming generality of the findings on spacing that have been reported in the literature, we suggest, on the basis of the arguments outlined above, that there may be conditions under which massing rather than spacing of practice should result in better learning. To date, though, no such reversal of the spacing effect has ever been reported.

We tested the hypothesis that there would be an interaction between study time and spacing in two experiments that were small variants of one another. In Experiment 4A, the study times were 1 s per item or 8 s per item; participants were given Spanish–English pairs at study and the Spanish words at test and were asked to type in the English translations. As will be seen, we obtained the predicted interaction between study time and spacing but a null rather than a negative spacing effect at the fast presentation rate. In an effort to obtain the predicted crossover, in Experiment 4B, we made the task slightly harder by presenting English–Spanish pairs at study and the English word at test, asking for the Spanish translation, and we also decreased the fast presentation rate to 0.5 s.

Method

Participants, design, and materials. Participants were Columbia University students who received partial course credit or pay for their participation. There were 24 participants in each of Experiments 4A and 4B.

The experiments each consisted of a 2 (spacing: massed or spaced) \times 2 (presentation time per instance: 1 s vs. 8 s in Experiment 4A and 0.5 s vs. 8 s in Experiment 4B) within-participants design with 24 word pairs per treatment combination. The order of the timing conditions—a blocked variable—was counterbalanced across participants. The materials were the 96 medium-difficulty pairs used in Experiment 3.

Procedure. Participants were told that they would be learning 96 English–Spanish vocabulary pairs. In Experiment 4A, they were told that they would be given the Spanish words at test and would be asked to type

in the English translations; in Experiment 4B, they were told that they would be given the English words at test and would be asked to type in the Spanish translations. The word pairs were presented in two blocks, one for each presentation time condition. In each block, after all of the pairs were presented, there was a distractor task during which participants counted backwards by 3s for 30 s, and then the cue words were presented in random order and the participants typed in their translations. Participants were told to guess during the test if they could not think of the exact answer. The second block started immediately after the first block ended.

Massing and spacing within each block. During presentation, each massed pair was presented twice in a row. An interstimulus interval of 0.1 s was included in each pair's presentation time, during which the screen was blank. Thus, for massed presentation in the 1-s condition, the pair was on screen for 0.9 s, the screen was blank for 0.1 s, and then the same pair was visible for another 0.9 s followed by another 0.1 s of blank screen. In the spaced condition, the two presentations of each pair were separated by seven other presentations, some of which were massed and some of which were spaced, as outlined below.

Primacy and recency effects, as well as lag-to-test effects, make the order in which word pairs are presented crucial (see Shaughnessy, Zimmerman, & Underwood, 1972, for further discussion of these problems). For example, if all the spaced pairs were presented at the start or the end of the list, they might be remembered better regardless of spacing because of serial position effects. To ensure that the average time from presentation to test was equivalent for the massed and spaced conditions, we devised a system in which pairs could be spaced moderately far apart but also grouped into different sets, so that all of the pairs (massed and spaced) in a given set were approximately equally distant from the test. The pairs were presented in sets of eight, with each of the eight pairs being presented twice. In each set, each spaced pair was separated by presentations of each of the three other spaced pairs and by two of the four massed pairs. For example, in one permutation, the pairs were presented in the following order: S1, S2, M1, M1, S3, S4, M2, M2, S1, S2, M3, M3, S3, S4, M4, M4 (where S1, M1, etc. represent spaced [S] and massed [M] pairs and each number identifies a particular pair). Six such sets of eight pairs were presented in each timing condition. We used different permutations for the six sets so that the participants would not recognize a pattern during presentation. Each permutation had the same structure; for example, another permutation was S1, S2, S3, M1, M1, S4, M2, M2, S1, S2, S3, M3, M3, S4, M4, M4. As a result of this ordering system, the average distance from the last presentation of a pair to the start of the test was very similar, 44 pairs for the spaced condition and 46 pairs for the massed condition. This small difference in lag to test was further offset by the distractor task.

Results

Experiment 4A. As is shown in Figure 6A, accuracy was better in the 8-s condition than in the 1-s condition, $F(1, 23) = 59.08$, $MSE = 0.03$, $ES = .72$. Accuracy was also better in the spaced condition than in the massed condition overall, $F(1, 23) = 6.51$, $MSE = 0.01$, $ES = .22$, but this was qualified by the predicted significant Presentation Time \times Spacing interaction, $F(1, 23) = 6.75$, $MSE = 0.01$, $ES = .23$. Planned comparisons showed that in the 8-s condition, accuracy was better in the spaced condition than in the massed condition, $t(23) = -3.07$, $p = .003$ (one-tailed). In the 1-s condition, there was no difference, $t(23) = 0.15$, $p = .440$ (one-tailed).

Experiment 4B. As is shown in Figure 6B, accuracy was better in the 8-s condition than in the 0.5-s condition, $F(1, 23) = 86.00$, $MSE = 0.03$, $ES = .79$. Accuracy was also better in the spaced condition than in the massed condition, $F(1, 23) = 9.95$, $MSE = 0.00$, $ES = .30$, but this main effect was qualified by the predicted significant Spacing \times Presentation Time interaction, $F(1, 23) =$

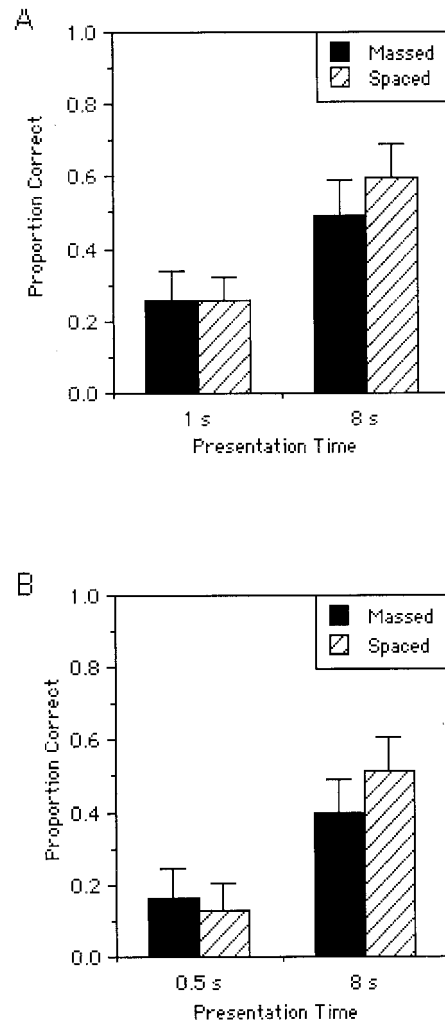


Figure 6. The interaction between spacing and presentation time in Experiment 4A (A) and Experiment 4B (B). Error bars indicate 95% confidence intervals.

23.80, $MSE = 0.01$, $ES = .51$. Planned comparisons showed that in the 8-s condition, accuracy was better in the spaced condition than in the massed condition, $t(23) = -4.95$, $p < .0001$ (one-tailed). In the 0.5-s condition, accuracy was better in the massed condition than in the spaced condition, $t(23) = 2.03$, $p = .027$ (one-tailed).

Discussion

These results conform well to the predictions inferred from the information uptake functions of Experiment 3. The results of both Experiments 4A and 4B showed the interaction between presentation time and spacing. Furthermore, in Experiment 4B, at the short (0.5 s) presentation time, it was disadvantageous to interrupt processing—massing resulted in reliably better performance than did spacing. When the presentation time was intermediate (1 s), there was no significant difference between massed and spaced practice. Only when presentation times were long (8 s) was there a difference favoring spaced practice.

Spaced practice has often been advocated as a way to optimize learning. A host of experiments have promoted this conclusion (e.g., Dempster, 1988; Reder & Anderson, 1982), and it has been obtained in many paradigms using many populations and, indeed, species. We do not challenge the conclusion that spaced practice is best when the information uptake on the first presentation is complete—that is, when little or no further learning would be obtained by continued immediate perseverance. However, the dynamics of the information uptake need to be taken into account, and the learning process should not be interrupted prematurely. The results of this experiment, then, present a strong caveat to the generality of the conclusion that one should always defer practice until later. If the learning has not plateaued on the first presentation, discontinuing study, by spacing practice, is liable to hurt performance as compared with continuing to study immediately.

Conclusion

How do the two models of study-time allocation outlined in the introduction fare given the current data? The discrepancy reduction model fails, both in terms of what people do and what they should do. In contrast to what the discrepancy reduction model predicts, in Experiment 1, people did not devote their time selectively to the most difficult items, nor should they have. The lack of increase in performance on the difficult items coupled with an order of magnitude increase in study time justifies Nelson and Leonesio's (1988) epitaph on these items of "labor in vain." Selective studying of the most difficult items also did not help performance in the second experiment. In fact, it produced the worst performance. We do not dispute that if people were given an unlimited amount of time and the goal were complete mastery of all of the items, they would need to devote the most time to the most difficult items. Indeed, this fact may constitute the best operational definition of difficulty. But that does not mean that this is the best thing to do with one's time in every situation. Under more realistic circumstances in which people have limited time and cannot reach perfection, they neither use nor should use the discrepancy reduction strategy.

The region of proximal learning framework fared better. The general pattern of the data fit well with the underlying assumptions about learning that are embodied in more general region of proximal learning theories (Atkinson, 1972a, 1972b; Berlyne, 1978; Hebb, 1949; Piaget, 1952; Vygotsky, 1987). These theorists have proposed that learning proceeds best (and, as noted by Berlyne, one's aesthetic sense is most rewarded) by attending to those concepts and events that are nearest to one's current understanding and only then, with the enriched conceptual base, going on to integrate items that are more difficult. This type of theory has no difficulty explaining most of the findings presented here. First, as was shown in Experiment 1 and by Thiede and Dunlosky (1999), people chose to study the easiest items first. However, according to the information uptake functions of Experiment 3, little time is required to master those easy items. Thus, it is well worth studying those items for a short time, but they are likely to be transferred rather quickly to a state where additional study time is ineffective. Once that occurs, the region of proximal learning theory indicates that the items that are next in line to being learned (i.e., the medium-difficulty items) should be studied. And, indeed, people turned to the items of medium difficulty next, as was shown in

Experiment 1. They studied the medium-difficulty items longer than they did the easy items—again, as is appropriate given the information uptake functions.

Perhaps the only puzzling aspect of the data on people's choice of what to study from the perspective of the region of proximal learning framework is why they studied the most difficult items at all. The answer might be that they had to at least check on those items to make the assessment that studying them was not beneficial. But checking takes time, and they could not do so without seeming to study them. In addition, because the participants had only three options, once they had studied the easy and medium-difficulty items, they had nowhere else to go but to the difficult items. If they had had other easy and medium-difficulty items to return to, they might have done so.

It is perhaps a truism that learning is a deeply dynamic process. However, current understanding of the relation between metacognition and study-time allocation has not taken these dynamics into account. It is abundantly clear, though, from the data presented in this article that a strong correlation between JOLs and study-time allocation—a result that is predicted by the discrepancy reduction model and that has been taken to mean that people's metacognitions control their study-time allocation—indicates what is very likely to be a suboptimal strategy rather than the reverse. Even the region of proximal learning framework, which is fundamentally dynamic and which was well-supported by the data presented here, will require refinement to allow assessment of the benefits of continuing (or declining) to study when factors that take up precious study time, such as the costs of checking, choosing, and switching, are taken into account. We suggest that only by coming to terms with the intrinsic dynamics of information uptake and how those dynamics interact with the individual's online monitoring of his or her own state of learning vis-à-vis a particular item under consideration at the time of monitoring, will psychologists be able to construct a more detailed and realistic theory of metacognitively guided study-time allocation. Such a theory should make smooth predictions not only about what items people should study, when they should decline to study them further, and when they should return to study them again, but also when and why they will encounter difficulties in making such decisions.

References

- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396–408.
- Atkinson, R. C. (1972a). Ingredients for a theory of instruction. *American Psychologist*, 27, 921–931.
- Atkinson, R. C. (1972b). Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology*, 96, 124–129.
- Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, 4, 316–321.
- Belmont, J. M., & Butterfield, E. C. (1971). Learning strategies as determinants of memory deficiencies. *Cognitive Psychology*, 2, 411–420.
- 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.
- Berlyne, D. E. (1978). Curiosity and learning. *Motivation and Emotion*, 2, 97–175.
- 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., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding. *Journal of Verbal Learning and Verbal Behavior*, 9, 567–572.
- Calfee, R. C., & Atkinson, R. C. (1965). Paired-associate models and the effects of list length. *Journal of Mathematical Psychology*, 2, 254–265.
- Carew, T. J., Pinsky, H. M., & Kandel, E. R. (1972, January 28). Long-term habituation of a defensive withdrawal reflex in *Aplysia*. *Science*, 175, 451–454.
- Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review*, 97, 19–35.
- Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.
- 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.
- Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist*, 43, 627–634.
- Dufresne, A., & Kobasigawa, A. (1988). Developmental differences in children's spontaneous allocation of study time. *Journal of Genetic Psychology*, 149, 87–92.
- Dufresne, A., & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: Differential and sufficient aspects. *Journal of Experimental Child Psychology*, 47, 274–296.
- Dunlosky, J., & 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. (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., & 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.
- Ebbinghaus, H. E. (1964). *Memory: A contribution to experimental psychology* (Henry A. Ruger & Clara E. Bussenius, Trans.). New York: Dover. (Original work published 1885)
- Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, 15, 1–16.
- Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, 7, 95–112.
- Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 371–377.
- Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.
- Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77–99). Hillsdale, NJ: Erlbaum.
- Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 10, pp. 47–91). New York: Academic Press.
- Hyland, M. E. (1988). Motivational control theory: An integrative framework. *Journal of Personality and Social Psychology*, 55, 642–651.
- Johnston, W. A., & Uhl, C. N. (1976). The contribution of encoding effort and variability to the spacing effect on free recall. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 153–160.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Kellas, G., & Butterfield, E. C. (1971). Effect of response requirement and type of material on acquisition and retention performance in short-term memory. *Journal of Experimental Psychology*, 88, 50–56.
- Kobasigawa, A., & Dufresne, A. (1992). *Differential allocation of study time by Grade 3 children*. Unpublished manuscript.
- Kobasigawa, A., & Metcalfe-Haggert, A. (1993). Spontaneous allocation of study time by first- and third-grade children in a simple memory task. *The Journal of Genetic Psychology*, 154, 223–235.
- Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition*, 9, 149–171.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103, 490–517.
- Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, 131, 147–162.
- Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London: Academic Press.
- Le Ny, J. F., Denhiere, G., & Le Taillanter, D. (1972). Regulation of study-time and interstimulus similarity in self-paced learning conditions. *Acta Psychologica*, 36, 280–289.
- Lord, R. G., & Hanges, P. J. (1987). A control system model of organizational motivation: Theoretical development and applied implications. *Behavioral Science*, 32, 161–178.
- Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, 8, 828–835.
- Masson, M. E. J., & Loftus, G. R. (in press). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*.
- 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.
- Mazzoni, G., Cornoldi, C., Tomat, L., & Vecchi, T. (1997). Remembering the grocery shopping list: A study on metacognitive biases. *Applied Cognitive Psychology*, 11, 253–267.
- Melton, A. W. (1967, October 27). Repetition and retrieval from memory. *Science*, 158, 532.
- Menzel, R., Manz, G., Menzel, R., & Greggers, U. (2001). Massed and spaced learning in honeybees: The role of CS, US, the intertrial interval, and the test interval. *Learning and Memory*, 8, 198–208.
- Metcalfe, J. (1998). Cognitive optimism: Self-deception or memory-based processing heuristics? *Personality and Social Psychological Review*, 2, 100–110.
- Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General*, 131, 349–363.
- Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science*, 2, 267–270.
- Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, 5, 207–213.
- Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 676–686.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework

- and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–141). New York: Academic Press.
- 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.
- Pashler, H. (2002). *Spacing versus error minimization: Effects on learning rate*. Paper presented at the 43rd Annual Meeting of the Psychonomic Society, Kansas City, MO.
- Pelegriña, S., Bajo, M. T., & Justicia, F. (2000). Differential allocation of study time: Incomplete compensation for the difficulty of the materials. *Memory, 8*, 377–392.
- Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.
- Powers, W. T. (1973). *Behavior: The control of perception*. Hawthorne, NY: Aldine de Gruyter.
- Reder, L. M., & Anderson, J. R. (1982). Effects of spacing and embellishment on memory for main points of a text. *Memory & Cognition, 10*, 97–102.
- Rumelhart, D., & McClelland, J. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press/Bradford Books.
- Russo, R., Mammarella, N., & Avons, S. E. (2002). Toward a unified account of spacing effects in explicit cued-memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 819–829.
- Shaughnessy, J. J., Zimmerman, J., & Underwood, B. J. (1972). Further evidence on the MP-DP effect in free recall learning. *Journal of Verbal Learning and Verbal Behavior, 11*, 1–12.
- 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.
- Stevens, S. S. (1962). The surprising simplicity of sensory metrics. *American Psychologist, 17*, 29–39.
- 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.
- Vygotsky, L. S. (1987). *The collected works of L. S. Vygotsky, Volume 1: Problems of general psychology* (R. W. Rieber & A. S. Carton, Eds.). New York: Plenum Press.
- Waugh, N. C. (1970). On the effective duration of a repeated word. *Journal of Verbal Learning and Verbal Behavior, 9*, 587–595.
- Whitten, W. B., & Bjork, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning and Verbal Behavior, 16*, 465–478.
- Zacks, R. T. (1969). Invariance of total learning time under different conditions of practice. *Journal of Experimental Psychology, 82*, 441–447.

Received May 8, 2002

Revision received June 3, 2003

Accepted June 12, 2003 ■

New Editors Appointed, 2005–2010

The Publications and Communications Board of the American Psychological Association announces the appointment of two new editors for 6-year terms beginning in 2005:

- *Journal of Consulting and Clinical Psychology*: **Annette M. La Greca**, PhD, ABPP, Professor of Psychology and Pediatrics, Department of Psychology, P.O. Box 249229, University of Miami, Coral Gables, FL 33124-0751.
- *Developmental Psychology*: **Cynthia García Coll**, PhD, Brown University, 21 Manning Walk, Providence, RI 02912.

Electronic manuscript submission. As of January 1, 2004, manuscripts should be submitted electronically via the journal's Manuscript Submission Portal. Authors who are unable to do so should correspond with the editor's office about alternatives. Portals are available at the following addresses:

- For *Journal of Consulting and Clinical Psychology*, submit via www.apa.org/journals/ccp.html.
- For *Developmental Psychology*, submit via www.apa.org/journals/dev.html.

Manuscript submission patterns make the precise date of completion of the 2004 volumes uncertain. Current editors, Mark B. Sobell, PhD, and James L. Dannemiller, PhD, respectively, will receive and consider manuscripts through December 31, 2003. Should 2004 volumes be completed before that date, manuscripts will be redirected to the new editors for consideration in 2005 volumes.