Journal of Cognitive Psychology, 2014
Vol. 26, No. 8, 928–942, http://dx.doi.org/10.1080/20445911.2014.967771

Routledge
Taylor & Francis Group

Is focusing on unknown items while studying a beneficial long-term strategy?

Nate Kornell[1] and Katrina E. Flanagan[2]

[1]Department of Psychology, Williams College, 18 Hoxsey Street, Williamstown, MA 01267, USA
[2]Harvard Law School, Boston, MA, USA

The amount students learn is influenced by the decisions they make while they study. Prior research shows that when people decide what to study (and what not to), they almost universally prioritise information they do not know over information they do know. We found that focusing on unknown items was beneficial when memory was tested immediately after studying, but not when memory was tested after a delay of one or two weeks. These findings support a model in which studying unknown items causes many items to rise above a recall threshold, which results in a short-term advantage, but also leaves the items vulnerable to forgetting. Studying known items, by contrast, moves items farther above threshold, such that they remain above threshold for longer. These findings have implications for students and teachers who want to maximise long-term learning.

Keywords: Delay; Learning; Memory; Metacognition; Studying.

Making appropriate decisions about what to study can have a meaningful impact on learning (for reviews, see Bjork, Dunlosky, & Kornell, 2013; Dunlosky & Ariel, 2011a; Nelson & Narens, 1994; Son & Kornell, 2008; Thiede, 1999; Thiede & Dunlosky, 1999). One of the most frequent decisions is what material to focus on and what to skip over. In laboratory research on this type of decision, a principle has emerged: People consistently put aside items that they have recalled on a previous trial, and concentrate on items they do not yet know (e.g., Karpicke, 2009; Kornell & Bjork, 2008; Metcalfe & Kornell, 2003, 2005; Son & Metcalfe, 2000).[1] This strategy, which we will refer to as focusing on unknown items, is also common outside of the laboratory.

Three influential models of study-time allocation predict that people will focus on unknown items (e.g., Cull & Zechmeister, 1994; Son, 2004;

[1] The main exception to this rule is situations in which the learner's goal is to learn only a subset of the items, which can make studying the best-known items an attractive strategy (Ariel et al., 2009; Thiede & Dunlosky, 1999). In addition, when there are too many unknown items to study in the allotted time, learners tend to prioritise unknown items that they are close to knowing over unknown items that they know less well (Metcalfe, 2002, 2009; Metcalfe & Kornell, 2003, 2005; Son & Metcalfe, 2000).

Correspondence should be addressed to Nate Kornell, Department of Psychology, Williams College, 18 Hoxsey Street, Williamstown, MA 01267, USA. E-mail: nkornell@gmail.com

A Scholar Award from the James S. McDonnell foundation supported this research, which was conducted as part of the second author's undergraduate senior thesis. Andrew Butler, Sean Kang and Katherine Rawson made valuable comments on this manuscript.

for reviews, see Dunlosky & Ariel, 2011a; Son & Metcalfe, 2000). The discrepancy reduction model claims that people focus on the items they know least well (e.g., Dunlosky & Hertzog, 1998), whereas the region of proximal learning model predicts that people focus on the easiest items that they do not already know (see Metcalfe, 2002, 2009; Metcalfe & Kornell, 2003, 2005; but see Dunlosky & Ariel, 2011b). The agenda-based regulation model claims that when deciding what to study, people are influenced by their goals (e.g., they study what is likely to be tested; e.g., Ariel, Dunlosky, & Bailey, 2009) and habits (e.g., readers of English are biased to move from left to right; e.g., Dunlosky & Ariel, 2011b). Because the research presented in this article concerns the relationship between item difficulty and study choices—and the point of the agenda-based regulation model is to examine cues other than item difficulty—we do not focus on agenda-based regulation model in what follows.

The discrepancy reduction and region of proximal learning models make claims about how people choose to study (e.g., they study items that they do not know). Both models also make a second type of claim: People *should* study according to the model, because doing so is effective. Past research has shown that in addition to being common, focusing on unknown items can be effective (e.g., Kornell & Metcalfe, 2006; Nelson, Dunlosky, Graf, & Narens, 1994).

There is a gap in the literature with respect to the effect of studying unknown items, however, because in prior studies memory was tested after a short retention interval (retention interval refers to the time that elapses after one has stopped studying but before one is tested). Although long-term learning is generally more important than short-term learning, it has yet to be examined. Our research sought to fill this gap. From a theoretical perspective, the main question was whether existing theories of study time allocation accurately capture the long-term effects of people's study strategies.

As explained earlier, previous evidence suggests that when allowed to decide which items to study, people will focus on unknown items. We made three predictions regarding this study strategy, which are explained in the sections that follow. First, focusing on unknown items is beneficial in the short term. Second, we predicted this benefit would decrease as retention interval increased (i.e., that there would be an interaction between study strategy and delay). Third, the study strategy people choose is unlikely to vary depending on the length of the delay between study and test. Combined, these predictions suggest that when people are allowed to select items for study, they will focus on unknown items and will benefit from doing so in the short term, but not in the long term.

FOCUSING ON UNKNOWN ITEMS IS BENEFICIAL IN THE SHORT TERM

As mentioned earlier, multiple studies have shown that people learn more when they focus on unknown items (e.g., Nelson et al., 1994). On the surface, then, it might seem there is a contradiction in the literature because there is also evidence that overlearning—studying items that one has already recalled—also causes people to learn more (e.g., Vaughn & Rawson, 2011, 2014). It might seem that overlearning is the opposite of focusing on unknown items. In fact, though, in most studies showing benefits of overlearning, the overlearning condition involves continuing to study instead of terminating study (i.e., more time is spent studying in the overlearning than the control condition). By this definition, overlearning is an effective way to learn (e.g., Vaughn & Rawson, 2014), and it is most effective when the restudy involves self-testing rather than presentation (Karpicke & Roediger III, 2008). A meta-analysis has shown that overlearning can be moderately effective across a variety of tasks (Driskell, Willis, & Copper, 1992).

In most research on study decisions, however, participants are asked to prioritise between items. That is, spending time on known items (i.e., overlearning) does not mean studying more; it means spending less time on unknown items. In this situation, as mentioned earlier, there is considerable evidence that studying unknown items is beneficial in the short term (e.g., Kornell & Metcalfe, 2006). Related evidence shows that overlearning is relatively inefficient compared to doing the same amount of studying spaced out across time (Rohrer, 2009).

The current state of the research, therefore, suggests that when forced to prioritise, people choose to study unknown items, and when they instead study known items (i.e., overlearn) they learn less as a result. Given this research, though, there is another ostensible contradiction in the literature. Research has shown that ceasing study after a single correct response can be counterproductive. In a study by Kornell and Bjork (2008), participants who were allowed to "drop" items from further study were most likely to stop studying after answering correctly once. Yet a subsequent experiment showed that participants learned more when forced to study each word pair at least once after answering it correctly—even though doing so took time away from studying unknown items—than when they were allowed to drop the item after one correct response.

One explanation of Kornell and Bjork's (2008) findings is that participants adopted an inappropriate norm of study (Le Ny, Denhiere, & Taillanter, 1972; Dunlosky & Thiede, 1998). Instead of studying until they had retrieved the item multiple times, they stopped after retrieving it once. In the next section, we explain why retrieving an item only once can be an ineffective strategy. In short, doing so does not create memories that are strong enough to withstand the ravages of time and forgetting.

LONG-TERM EFFECTS OF FOCUSING ON UNKNOWN ITEMS

We predicted an interaction such that focusing on unknown items would be effective in the short term but not in the long term. This prediction follows from a threshold model of memory that takes into account differences in memory strength between items (e.g., Kornell, Bjork, & Garcia, 2011; Nelson, 1993). At any given moment, a memory has a level of "retrieval strength" that determines whether or not it can be retrieved (e.g., Bjork & Bjork, 1992). If retrieval strength exceeds a certain threshold, the item can be retrieved (that is, it can be considered known, at least in the short term).

Focusing on unknown items, translated into these terms, means that people choose to study items that are below threshold rather than items that are already above threshold. The common strategy of studying below-threshold items until they are just above threshold, and then not studying them anymore, leads to a large number of memories being above threshold, but not by very much (see Figure 1). This strategy will lead to good performance on a test that takes place after a short delay, when these items are still above threshold. However, as time passes, retrieval strength decreases (i.e., people forget). Items that were slightly above threshold immediately after study will tend to move below threshold after a longer delay (Figure 1). A better strategy to maximise long-term learning (i.e., performance on a delayed test) might be to study known items, which insures that they move far enough above threshold that they are still recallable on a delayed test.

Figure 1 presents hypothetical data that illustrate how two study strategies—the common strategy of focusing on unknown items and the uncommon strategy of studying known items—might affect performance on short-term and long-term tests. Based on Figure 1, we predicted an interaction: Focusing on unknown items should be beneficial in the short term, replicating previous studies, but the advantage of studying unknown items should shrink as the retention interval grows.[2]

WILL EXPECTED RETENTION INTERVAL AFFECT STUDY DECISIONS?

If focusing on unknown items is beneficial for short retention intervals but not long retention intervals, then to study effectively, people should take retention interval into account when making study decisions. To investigate whether they do so, participants in Experiment 3 were allowed to choose how they studied in preparation for a test that they knew would occur after either a short retention interval or a long retention interval.

We predicted that participants' expectations about the retention interval would not affect their study choices. Prior research on the judgements of learning—judgements of the probability that one will recall information on a later test—have consistently exhibited a pattern of data that has been referred to as a stability bias (Kornell & Bjork, 2009): In general, people do not seem to take future forgetting into account when making judgements about their own learning (Carroll, Nelson, & Kirwan, 1997; Koriat, Bjork, Sheffer, & Bar, 2004). Instead, they act as though their knowledge will be the same in the future as it is now. If people are insensitive to the effects of retention interval on knowledge, it is unlikely that they will study differently depending on retention interval.

[2] An important caveat needs to be stated here. Our predictions concern the percentage of items that can be recalled on a test, which is essentially the percentage of items above threshold. The percentage of items that can be recalled on a test is not the same as memory strength, and in fact our point is that changes in memory strength are not necessarily reflected in the percentage of items that can be recalled (see Kornell et al., 2011). In other words, as we explain in the general discussion, our predictions have to do with recall, not underlying memory strength.

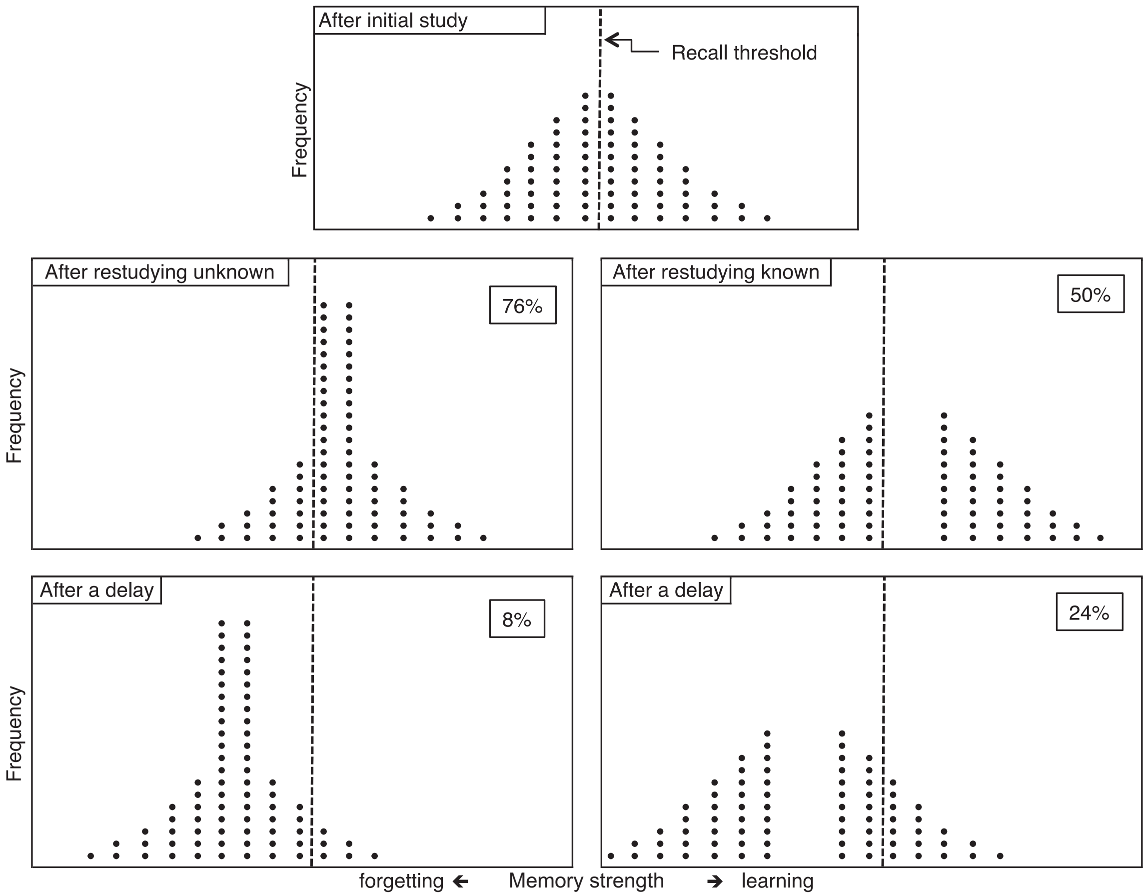

Figure 1. Data from a hypothetical experiment. After initial study (top), items are normally distributed around a recall threshold. Items that are restudied (middle panels) get stronger (i.e., move right). Restudying unknown items (middle panel, left) causes items to accumulate just above threshold because studied items move above threshold and join unstudied items. Studying items that were already above threshold (middle panel, right) moves them farther above threshold. In the short term, more items are above threshold after studying unknown (76%) than known items (50%). After a delay (bottom panels), however, this balance reverses (8% versus 24%). In this hypothetical scenario, changes in the distribution of item strength create the appearance of differential forgetting rates between the conditions on the left and right, but it is important to note that the amount of strength gained from studying, and lost due to forgetting, is equal in the left and right columns.

Moreover, in the present study recognising the existence of forgetting would merely be the first step. If our first prediction is correct, and the rate of apparent forgetting is greater when people focus on unknown items than it is when they focus on known items, then participants would have to foresee not just that they forget, but that they forget *at different rates* depending on how they study, which seemed quite unlikely. Previous studies have shown that people are insensitive to the degree to which rates of apparent forgetting will differ between conditions, for example as a function of being tested versus restudying (Roediger III & Karpicke, 2006a). Thus, we predicted that participants would not adjust their study choices based on the expected retention interval.

We tested the three predictions described in this introduction in three experiments. In Experiments 1 and 2, participants studied word pairs and then took an initial test. There was a subsequent study phase, during which participants focused on known items or unknown items. A final test took place after either a few min or multiple days. The procedure was similar in Experiment 3, except that the participants were asked to decide which items they wanted to study.

EXPERIMENT 1

Participants in Experiment 1 studied 60 word pairs, took an initial test and then restudied half of the pairs using test-with-feedback trials.[3] There were two restudy conditions. In the restudy known condition, participants restudied items that they had recalled on the initial test (but not items that they had not recalled). In the restudy unknown condition, they restudied items they had not recalled (but not items that they had recalled). Participants then took a final test after either three min or one week.

Method

Participants. Participants were recruited using Amazon's Mechanical Turk, a Web site on which users complete tasks for pay. Mechanical Turk provides access to a group of participants that is more diverse in terms of age and occupation than the typical lab experiment's college student pool, and recent studies have shown Mechanical Turk results that replicate laboratory findings (e.g., Mason & Suri, 2012). In addition, we believe the online platform used for this study is a better approximation of many modern learning experiences—which are increasingly taking place over the Internet—than an in-lab experiment.

To avoid subject selection effects, all participants—even those in the immediate test condition —were required to complete both sessions to be included in the data analysis. Twenty-two participants completed the experiment. Additional participants were excluded for various reasons: Six did not complete all of Session 1, eight did not complete Session 2 and three reported that they were familiar with the experimental materials. Of the 22 participants included in the analysis, 18 were female, and ages ranged from 19 to 62. All participants lived in the USA and were fluent in English. Eleven participants were randomly assigned to the immediate test condition and 11 to the delayed test condition.

Session 1 took about 25 min and Session 2 took about five min. Participants were paid $1.50 for completing Session 1 and another $1.00 for completing Session 2.

Materials. The stimuli were 60 word pairs (e.g., firm-grip; agony-defeat). According to norms reported by Nelson, McEvoy, and Schreiber (1998), each pair had a forward association strength of .10 or .11, which means that when people were shown the first word in each pair (e.g., firm) and asked to respond with the first word that came to mind, 10–11% of people produced the second word in the pair (e.g., grip).

Design. The experiment had a 2 × 2 mixed design. Test delay was manipulated between participants; after the study phase ended, participants waited either three min or one week to take the final test.

Study strategy was manipulated within participants. For each participant, some items were assigned to the restudy known condition and others were assigned to the restudy unknown condition. The restudy known and restudy unknown conditions were created after the experiment; the creation of these two conditions was based on participants' accuracy on the initial test. Items assigned to the restudy known condition were presented for restudy if and only if they had been recalled successfully on the initial test. Items assigned to the restudy unknown condition were presented for restudy if and only if they had not been recalled on the initial test. (For more information about how items were assigned to the restudy known and restudy unknown conditions, see the results section and Table 1.)

Procedure. The experiment was conducted online. There were four stages: Initial study, initial test, restudy and final test. During the initial study phase, word pairs were presented one at a time, in a random order and participants studied each word pair once. The cue and target of a given pair were shown on the screen until the participant pressed a button to move on to the next pair.

After studying all of the pairs, participants proceeded to the initial test phase, during which each cue word was presented (in a random order different from the initial study phase), and participants were asked to type in the target. They were not given feedback about their responses.

After the initial test ended the restudy phase began. Half of the pairs were randomly selected for restudy, and participants restudied these pairs twice each. They restudied each of the pairs in random order, and then they restudied each one again in a new random order. Restudy trials consisted of tests with feedback: Participants

[3] In the literature on retrieval effects, the term "study" is often reserved for trials on which participants were not tested. In normal parlance, however, reading and retrieval practice are both considered forms of studying. In this article, we use the terms "study" and "restudy" to refer to learning activities the participant engaged in, regardless of whether they involved presentation trials or retrieval trials.

TABLE 1
Mean accuracy on the final test (standard deviation in parentheses). The data from Experiment 3 are broken down in two ways

	Known/restudied	*Known/no restudy*	*Unknown/restudied*	*Unknown/no restudy*
Experiment 1				
Immediate	.976 (.044)	.922 (.041)	.885 (.159)	.137 (.188)
Delayed	.545 (.187)	.409 (.080)	.303 (.193)	.042 (.055)
Experiment 2				
Immediate	.955 (.069)	.659 (.236)	.819 (.160)	.193 (.287)
Delayed	.327 (.242)	.103 (.097)	.261 (.180)	.062 (.082)
Experiment 3				
Immediate	.990 (.037)	.920 (.115)	.795 (.301)	.036 (.082)
Delayed	.350 (.305)	.198 (.177)	.110 (.128)	.024 (.051)
	Chose no restudy/dishonor	Chose no restudy/honor	Chose restudy/honor	Chose restudy/dishonor
Experiment 3				
Immediate	.882 (.193)	.762 (.278)	.723 (.310)	.302 (.359)
Delayed	.244 (.217)	.170 (.144)	.128 (.141)	.027 (.053)

were presented with the first word in each pair and asked to type in the second word. They were given as much time as needed to enter an answer. Once they pressed enter to indicate that they were finished with the trial, the correct answer was displayed on the screen for 2 s, and then the next trial began. (Items were assigned to the restudy known condition or the restudy unknown condition after the procedure had ended; items were assigned to the restudy known condition if they were randomly selected to be studied and had been answered correctly or if they were randomly selected not to be studied and had been answered incorrectly; otherwise they were assigned to the restudy unknown condition.)

After the restudy phase, Session 1 ended for participants in the delayed test condition. Participants in the immediate test condition played the video game Asteroids for three min as a distractor task and then took a test (i.e., were shown each pair one at a time and asked to type in the target) without feedback on all 60 word pairs. All participants were emailed one week later and asked to take the delayed test on all of the word pairs. The delayed test was the same as the immediate test.

Results and discussion

Participants took an initial test before they restudied. Accuracy on the initial test in the immediate condition (M = .658, SD = .168) did not differ significantly from accuracy in the delayed condition (M = .613, SD = .139), $t(20)$ = .69, p = .498, d = .29. This lack of difference was expected because the test occurred before the delay manipulation.

Table 1 displays the proportion correct on the final test for each of four types of items. For the primary analyses the four types of items were recombined to create two conditions. The *restudy known* condition was made up of items from the first and last column of Table 1. The *restudy unknown* condition was made up of items in the second and third columns.

The means in the restudy known and restudy unknown conditions can be computed in two ways. One way, weighting by item, is to combine all items in a given condition and compute a mean for that condition. The other way, weighting by type, is to compute four means for each participant, corresponding to the four means in Table 1 and then average two means for the restudy known condition and the other two means for the restudy unknown condition. In conducting the primary analyses, we chose the latter approach, weighting by type, for two reasons. First, it is not affected by the number of items participants recalled on the initial test, which makes the results generalizable to more study situations. Second, the differences in the apparent rate of forgetting that we report are greater when weighted by item, so weighting by type is more conservative with respect to our hypotheses. Analyses weighted by item are presented in the Appendix.

As Figure 2 shows, the restudy unknown condition produced better overall accuracy on the final test than did the restudy known condition, $F(1, 20)$ = 65.69, p < .0001, η_p^2 = .77. Not surprisingly, participants also recalled more items on the immediate test than on the delayed test, $F(1, 20)$ = 138.83, p < .0001, η_p^2 = .87. Moreover, as Figure 2 shows, there was a significant interaction between these variables, $F(1, 20)$ = 31.78, p < .0001, η_p^2 = .61.

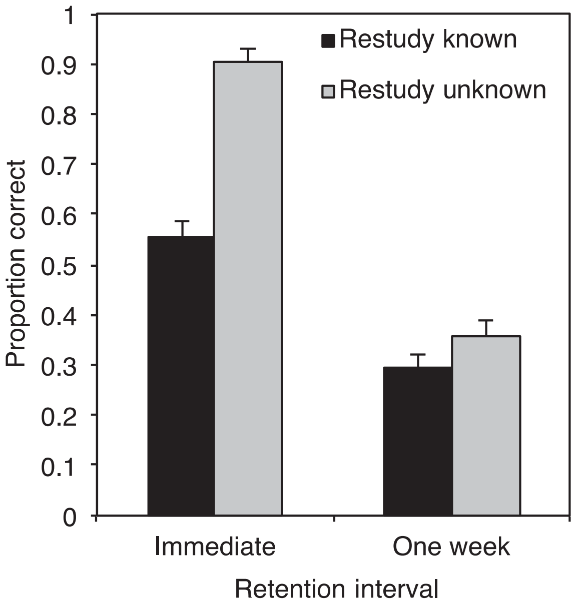

Figure 2. Proportion correct on the final test in Experiment 1. The data are split into conditions based on whether participants studied only items they had answered correctly during the initial test (restudy known) or only items they answered incorrectly (restudy unknown). Error bars represent 1 SEM.

Consistent with our prediction, restudying unknown items was highly beneficial when memory was tested immediately, but this benefit shrank as the retention interval increased. Because the results that are most interesting from an educational perspective have to do with long-term learning, we conducted a planned comparison of the means in the delayed condition. The restudy known (M = .294) and restudy unknown conditions (M = .356) did not differ significantly, $t(10) = -1.57$, p = .147, d = .47.

In Experiment 1, the study strategy that most students prefer—study what you do not know—enhanced learning, but only on an immediate test. Participants who were tested a week after they had finished studying did not benefit significantly from focusing on unknown items. This result calls into question the relevance to long-term learning of previous studies that have demonstrated benefits of focusing on unknown items (e.g., Kornell & Metcalfe, 2006; Nelson et al., 1994). The findings are consistent with the idea that restudying unknown items can move many items above the threshold for recall, but because they do not get very far above threshold, those items were likely to be forgotten relatively quickly (test accuracy fell from 90.4% to 35.6% between the immediate and delayed tests). The strategy of restudying known items, on the other hand, moves fewer items above threshold, but it also moves them far enough above threshold that fewer are forgotten (i.e., move from above to below threshold) a week later (test accuracy fell from 55.7% to 29.4% between the immediate and delayed tests).

EXPERIMENT 2

In Experiment 1, the apparent benefit of choosing to study unknown items decreased when the retention interval was increased from a few min to one week. We predicted that if we increased the retention interval further this trend would continue, and that a long enough retention interval might eliminate, or even reverse, the benefit of choosing to study unknown items. Experiment 2 differed from Experiment 1 in two important respects. The retention interval was increased from one week to two week, and half of the word pairs from Experiment 1 were replaced with more difficult pairs.

Method

Participants. Thirty-two participants completed the experiment. Additional participants were excluded for various reasons: 15 did not complete all of Session 1, 6 did not complete Session 2 and 1 reported that he or she was familiar with the experimental materials. Of the 32 participants included in the analysis, 22 were female and ages ranged from 19 to 58. Fifteen of the participants were randomly assigned to the delayed condition, and the other 17 were in the immediate condition. All participants were fluent English speakers who lived in the USA. They were paid $1.00 after completing each session.

Materials. The stimuli were 60 word pairs, half of which were easy (e.g., firm-grip) and half of which were difficult (e.g., meek-clothes). The easy pairs, which had a forward association strength of .10, were taken from Experiment 1. The difficult pairs also came from Experiment 1, but instead of using the intact pairs, which had a forward association strength of .11, we converted the pairs into non-associates by shuffling the targets and matching them to cues randomly. The change in materials was designed to create a larger difficulty gap between the known and unknown items. Item difficulty was not a variable of interest in our experimental design, however, and thus is not reported in the results.

Design and procedure. The design and procedure were the same as in Experiment 1. The only change was in the length of time between Session 1 and 2 in the delayed condition, which increased from one to two weeks.

Results and discussion

The results of Experiment 2 (Figure 3) replicated Experiment 1. Accuracy on the initial test in the immediate condition (M = .560, SD = .200) did not differ significantly from accuracy in the delayed condition (M = .494, SD = .236), $t(30)$ = .86, p = .397, d = .30. There were significant main effects of both study strategy and test delay, $F(1, 30) = 11.34$, $p = .002$, $\eta^2_p = .28$ and $F(1, 30) = 115.31$ $p < .0001$, $\eta^2_p = .79$, respectively. Moreover, there was a significant interaction, $F(1, 30) = 15.21$, $p = .0005$, $\eta^2_p = .34$, with the restudy unknown condition outperforming the restudy known condition on the immediate test but not on the delayed test. A paired samples t-test did not show a significant difference between performance in the restudy known (M = .195, SD = .129) and restudy unknown (M = .182, SD = .125) conditions on the delayed test, $t(14) = .41$, $p = .686$, $d = .11$.

These findings reinforce the findings from Experiment 1: Restudying unknown information does not necessarily enhance performance on a delayed test, even if it helps in the short term. The restudy unknown strategy produced more forgetting (a 57% point drop in accuracy) than the restudy known strategy (a 37% point drop in accuracy), resulting in equivalent performance on a test that occurred two weeks after participants had finished studying. Our hypothesis that after two weeks the restudy known condition might produce superior performance was not supported, however.

EXPERIMENT 3

Participants in Experiment 3 were asked to choose items for restudy. This experiment was important for understanding the long-term consequences of learners' decisions, because it is by choosing items to study that learners would be expected to enhance (or impair) their own learning. Experiment 3 resembled Experiments 1 and 2. The major

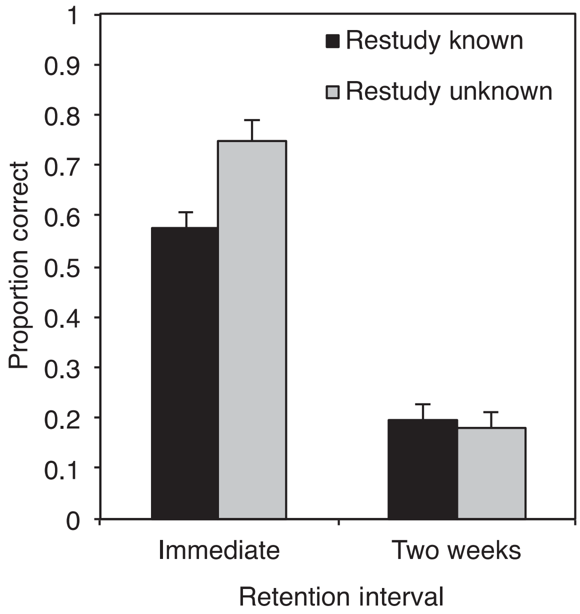

Figure 3. Proportion correct on the final test in Experiment 2. The data are split into conditions based on whether participants studied only items they had answered correctly on the initial test (restudy known) or only items they answered incorrectly (restudy unknown). Error bars represent 1 SEM.

methodological difference was that participants were asked to decide which items they did, and did not, want to restudy. Half of the study choices were honoured and half were dishonoured.

We made three predictions. First, participants would focus on unknown items. Second, they would do so regardless of the expected retention interval before they were to be tested. Third, focusing on unknown items would be effective in the short term but not in the long term.

The first two predictions are consistent with the research described in the introduction. In addition to being consistent with Experiments 1 and 2, the third prediction is also consistent with a recent study by Tullis and Benjamin (2012), who asked young and older adults to choose the items they wanted to study. Tullis and Benjamin found that both groups focused on unknown items. They also manipulated whether the participants' study choices were honoured (i.e., they studied only the items they had asked to study) or dishonoured (i.e., they studied only the items they had not asked to study). Tullis and Benjamin found that younger adults recalled more when their choices were honoured, but the older adults recalled *less*. That is, the older adults did better when they were forced (against their will) to study items they had recalled previously. Although Tullis and Benjamin did not manipulate delay, we reasoned that because older adults tend to forget more quickly than younger adults, the older adults might have forgotten more by the time of the final test than the younger adults did, almost as if the older adults experienced a longer delay. Although this idea is speculative, it is consistent with the prediction that the benefits of focusing on unknown items might be short-lived.

Method

Participants. Twenty-nine participants completed the experiment. Additional participants were excluded for various reasons: 20 did not complete all of Session 1, 37 did not complete Session 2 and 4 reported that they were familiar with the experimental materials. (The high number of participants who did not complete the experiments, particularly experiment 3, is a source of concern that will be addressed in the general discussion). Seventeen of the remaining participants were female, and ages ranged from 18 to 60 years old. All participants reported living in the USA and being fluent in English. Part one of the experiment took about 15 min, and part two took about five min. Participants were paid $1.00 after completing each session.

Materials. Thirty-two of the 60 word pairs used in Experiment 2 were randomly selected for use in Experiment 3. The number of pairs was decreased because a pilot study suggested that participants had difficulty keeping track of study decisions across 60 word pairs.

Design. The design of this experiment was a 2×2 mixed design, with test delay as a between-participants factor and whether a given restudy choice was honoured or dishonoured as a within-participants factor.

Procedure. In the previous experiments there were four stages: Initial study, initial test, restudy and final test. In Experiment 3 a study choice phase was added after the initial test. During the study choice phase, all 32 cue-words were displayed simultaneously. Items were displayed in different colours depending on whether the item had been answered correctly (black text) or incorrectly (red text) on the initial test. The colour cues were provided to help participants make informed study decisions. Participants were told what the colour cues signified but were given no instructions about which items to select for restudy. They were simply instructed to choose half of the items for restudy. They could not proceed to the next phase until they had chosen exactly half of the cues. Once they had done so, they pressed a button to proceed to the restudy phase.

During the selection phase, it was important that participants know when they would be tested, in case they wanted to choose their study strategy accordingly. Thus, during the initial instructions, participants in the delayed condition were told, "Finally, we might ask you to take a test two weeks from now on all of the word pairs you studied in the experiment". Participants in the immediate condition were told, "Next, you will play a video game for three min. You will then be tested on all of the word pairs you learned in the experiment. We might also ask you to participate in a follow-up experiment at a later date".

In the restudy phase, half of the participant's study choices were honoured and half were dishonoured. After the restudy phase, participants in the immediate condition took a final test and all participants took a test after a two-week delay, as in Experiment 2.

Results and discussion

The first question was whether participants chose to study unknown items. The answer is a resounding yes. They did not study all of the items they did not know, or only items they did not know, but only because they had to study exactly half of the items. For example, if a participant recalled exactly half of the items on the initial test, he could select every item he did not recall and no others. But if another participant answer three more than half correctly, the choice phase forced her to select at least three items that she had answered correctly, even if she was a strict disciple of the "focus on unknown items" school of studying. Similarly, if she recalled fewer than half of the items, she had the ability to select only unknown items, but she could not choose to study all of her unknown items. In short, unless a participant answered exactly half of the items correctly on the initial test, it was impossible for him or her to choose all of the unknown items, but no others, for restudy.

When this constraint is taken into account, the data suggest that participants in Experiment 3 followed the restudy unknown strategy without

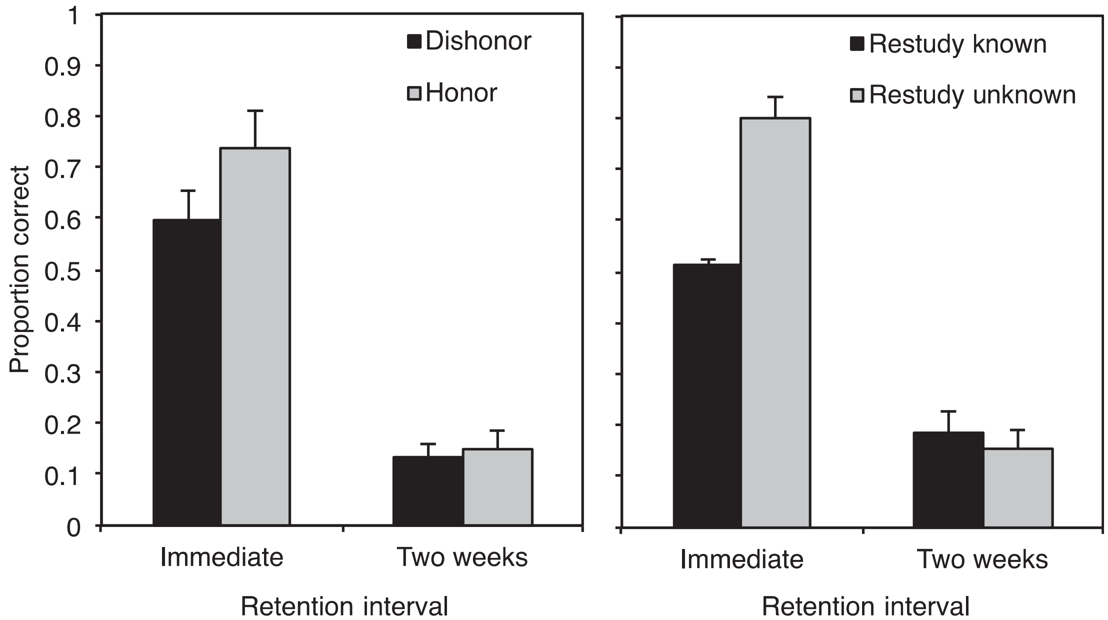

Figure 4. Proportion correct on the final test in Experiment 3. In the left panel, the data are split into conditions based on whether participants studied only the items they asked to study (honour) or only the items they did not ask to study (dishonour). In the right panel, the data are split into conditions based on whether participants studied only items they had answered correctly on the initial test (restudy known) or only items they had answered incorrectly (restudy unknown), in the same way as in Experiments 1 and 2 (Figures 2 and 3). Error bars represent 1 SEM.

exception. Participants who recalled more than half of the items universally chose to restudy all of the items they got wrong plus a few that they got right. All other participants could, and did, choose to study unknown items exclusively. In other words, no participant chose to study an item that they had answered correctly instead of an item they had not answered correctly. Given the utter dominance of this strategy, one might expect it to be highly effective.

We hypothesised, though, that honouring study choices would be effective on a short-term test, but that it might be less effective on a long-term test if participants tended to study items that they did not yet know. This hypothesis was supported (see Figure 4, left panel) by a significant interaction between test delay and whether or not study choices were honoured, $F(1, 27) = 4.28$, $p = .048$, $\eta_p^2 = .14$. There were also significant main effects of both test delay and honour/dishonour, $F(1, 27) = 59.48$, $p < .0001$, $\eta_p^2 = .69$ and $F(1, 27) = 6.30$, $p = .018$, $\eta_p^2 = .19$, respectively. A paired samples *t*-test did not show a significant difference between performance in the dishonour ($M = .135$) and honour ($M = .149$) conditions on the delayed test, $t(13) = .31$, $p = .763$, $d = .08$. As observed in the first two experiments, accuracy on the initial test in the immediate condition ($M = .546$, $SD = .283$) did not differ significantly from accuracy in the delayed condition ($M = .420$, $SD = .189$), $t(27) = 1.40$, $p = .173$, $d = .52$.

Because half of the participants' choices were honoured and half were not, we were able to repeat the analysis that we conducted in Experiments 1 and 2 by comparing restudy known versus restudy unknown items (see Figure 4, right panel). This secondary analysis produced a significant interaction, $F(1, 25) = 16.46$, $p = .0004$, $\eta_p^2 = .40$. There was also a significant main effect of study strategy and delay, $F(1, 25) = 10.32$, $p = .004$, $\eta_p^2 = .29$ and $F(1, 25) = 214.53$, $p < .0001$, $\eta_p^2 = .90$, respectively. A paired samples *t*-test did not show a significant difference between performance in the restudy known ($M = .187$ and restudy unknown ($M = .154$) conditions on the delayed test, $t(13) = .57$, $p = .576$, $d = .15$.

GENERAL DISCUSSION

Virtually everyone who studies prefers to focus on unknown items rather than known items. This strategy enhances short-term learning, but Experiments 1–3 showed that it was less effective for long-term learning. This result, which suggests that previous studies may have overestimated the benefit of studying unknown items (e.g., Kornell & Metcalfe, 2006; Nelson et al., 1994), fits with other findings showing that study strategies that enhance short-term performance do not always enhance long-term learning (e.g., Bjork, 1994; Bjork & Bjork, 2011; Simon & Bjork, 2001; Roediger III & Karpicke, 2006b). To be clear, studying unknown items did not significantly impair learning in any of the experiments.

Experiment 3 showed that participants placed a high priority on studying unknown items—in fact, they never chose to study a known item when they could choose an unknown item instead—but they did not benefit from this strategy on a test that occurred two weeks after they had finished studying. Why this bias was so strong is a topic for future research, but it is possible that participants had two motivations; first, they might have believed that studying unknown items is more productive than studying known items, and second, it might be more rewarding to study an unknown item than a known item because of the potential sense of accomplishment.

Theories of study time allocation

The fact that the study choices made by our participants did not enhance learning may have implications for two important theories of study time allocation, the discrepancy reduction model and the region of proximal learning model. Both models successfully predict the finding, from Experiment 3, that participants focused on unknown items. Neither theory predicted a difference in study choices as a function of retention interval, and consistent with the theories, the data showed no such difference.

The region of proximal learning model claims that not only do people study unknown items, they also benefit from doing so. As Kornell and Metcalfe put it, "(The region of proximal learning model] 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". (p. 621) Research on discrepancy reduction has drawn similar conclusions: Based on research showing that people benefited from choosing unknown items, Nelson et al. (1994) concluded that their research "demonstrates that people can use their metacognitions to allocate their restudy effectively". To the degree that these theories predict people benefit from adhering to them, the present findings are problematic because they suggest that this benefit is short-lived.

In both models, however, it might be that the root of the problem is not the strategy of studying unknown items. Instead, the problem might be with how people intuitively define "known". People tend to stop studying an item after they have recalled it correctly once (Karpicke, 2009; Kornell & Bjork, 2008), which might be an inappropriate norm of study (Le Ny et al., 1972). Perhaps people do not need to revise their strategies, but rather they need to understand that recalling something once does not make it "known". Long-term learning requires continuing to study after an item has been recalled. The best way of doing so is probably by testing oneself (Karpicke & Roediger III, 2008) and by using spaced practice (Rohrer, 2009).

Limitations

The learning materials used in this study were limited to a small collection of word pairs. It is not clear whether the results would have been the same with different materials. There is no obvious reason to predict that the results would have been different, but it is an empirical question.

The fact that a large number of participants did not complete the studies may be a cause for concern. Perhaps most worrisome are those who completed the first session but did not complete the second session (of which there were 8, 6 and 37 in Experiments 1–3, respectively). It is encouraging that the worst offender, Experiment 3, produced results that matched the results of Experiments 1 and 2. Nevertheless, if the participants who did not complete the study differed, in some meaningful way, from the participants who did complete the study, it might limit the generality of our findings. This problem would be especially troubling if some conditions in our studies were affected more than others, because then the findings would be systematically biased. There is no obvious reason to expect that dropouts biased our findings in a systematic way, especially because all participants had to complete both sessions to be included in the experiments. Still the high rate of dropouts may represent a weakness of the research presented here.

Memory strength versus test performance

Recall performance is essentially a measure of the percentage of items above threshold. Underlying memory strength, which cannot be measured directly, is not the same as recall. For example, if two memories are below threshold, neither can be recalled, but a memory that is just below threshold can be learned (i.e., moved above threshold) relatively easily, compared to an item that is far below threshold. Similarly, if two items are above threshold, the one that is far above threshold will remain recallable for longer than the one that is near threshold. Recall is all that matters at any given moment, but over time, underlying memory strength is probably more important than momentary recall because of the role it plays in subsequent learning and forgetting.

Our data show that there is a relative shift in *recall* over time whereby the benefit of studying unknown items tends to shrink. In terms of underlying memory strength, though, our data show no such shift. We did not measure underlying memory strength directly because no one has yet discovered a way to do so.

We do not claim that focusing on unknown items is better, or worse, than focusing on known items. Which one is better depends on whether the items will be studied later and when they will need to be retrieved. In fact, when we created Figure 1 we assumed that studying known and unknown items produced equivalent increase in memory strength (i.e., learning), but it is probable that studying unknown items actually produces more underlying learning than does studying known items (Bjork & Bjork, 1992), which would argue in favour of the popular strategy of studying unknown items—although this probability in no way takes away from the fact that recall can shift over time in a way that favours studying known items.

In short, Experiments 1–3 showed that the relative advantage of focusing on unknown items shifted over time, but this shift should not be misconstrued as a shift in underlying memory strength. That said, recall must be measured at some point in time. Given the long-term goals of education, we argue that recall is more interesting when it is measured after a meaningful delay than when it is measured immediately.

Practical implications

Students almost universally prioritise unknown items over known items. Doing so has clear benefits, especially if one's goal is to do well on a test tomorrow or sometime in the short term. The novel implication of the present studies, though, is that if one's goal is long-term learning, studying known items can also be effective, at least if such studying takes the form of retrieval practice followed by feedback.

The short-term benefit of studying unknown items can be quite apparent to learners, whereas the long-term benefits are difficult to perceive while one is studying. If student learning were assessed based on tests that took place years after studying, rather than at the end of the week or school year, the importance of encoding information in a way that really sticks might become more apparent. In the current education system, a demoralizingly large amount of knowledge that was once learned seems to fade into obscurity (Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996).

The present research is consequential, in part, because of the way students study and teachers teach. For example, the world's most common study strategy—cramming—is a campaign to get information just far enough above threshold that it can be remembered until the upcoming exam is over. This strategy obviously involves focusing on unknown items; studying information that one will know anyway, on the upcoming test, is seen as a waste of time. If the information being studied never makes it very far above threshold it might help explain why students can do well on a test and yet remember so little a month later. Teachers, too, are often tempted by strategies with short-term benefits. For example, instead of covering one topic in depth—thereby moving a relatively small amount of information far above threshold—teachers often prefer to teach more material in the same amount of time, thereby moving more information above threshold, but not moving it as far above threshold (e.g., Valverde & Schmidt, 1998; also see Reder & Anderson, 1982).

Original manuscript received August 2013
Revised manuscript received September 2014
Revised manuscript accepted September 2014
First published online October 2014

REFERENCES

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General*, *138*, 432–447. doi:10.1037/a0015928

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York, NY: Worth.

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

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In S. M. Kosslyn & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale, NJ: Routledge.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444. doi:10.1146/annurev-psych-113011-143823

Carroll, M., Nelson, T. O., & Kirwan, A. (1997). Tradeoff of semantic relatedness and degree of overlearning: Differential effects on metamemory and on long-term retention. *Acta Psychologica*, *95*, 239–253. doi:10.1016/S0001-6918(96)00040-6

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, *66*, 227–268. doi:10.3102/00346543066003227

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. doi:10.3758/BF03208896

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology*, *77*, 615–622. doi:10.1037/0021-9010.77.5.615

Dunlosky, J., & Ariel, R. (2011a). Self-regulated learning and the allocation of study time. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (54th ed., pp. 103–140). San Diego, CA: Elsevier. doi:10.1016/B978-0-12-385527-5.00004-8

Dunlosky, J., & Ariel, R. (2011b). The influence of agenda-based and habitual processes on item selection during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 899–912. doi:10.1037/a0023064

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: Lawrence Erlbaum Associates.

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*(1), 37–56. doi:10.1016/S0001-6918(97)00051-6

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, *138*, 469–486. doi:10.1037/a0017341

Karpicke, J. D., & Roediger III, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966–968. doi:10.1126/science.1152408

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*, 643–656. doi:10.1037/0096-3445.133.4.643

Kornell, N., & Bjork, R. A. (2008). Optimising self-regulated study: The benefits—and costs—of dropping flashcards. *Memory*, *16*, 125–136. doi:10.1080/09658210701763899

Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General*, *138*, 449–468. doi:10.1037/a0017350

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*, 85–97. doi:10.1016/j.jml.2011.04.002

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 609–622. doi:10.1037/0278-7393.32.3.609

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. doi:10.1016/0001-6918(72)90012-1

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's mechanical turk. *Behavior Research Methods*, *44*(1), 1–23. doi:10.3758/s13428-011-0124-6

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General*, *131*, 349–363. doi:10.1037/0096-3445.131.3.349

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*, 159–163. doi:10.1111/j.1467-8721.2009.01628.x

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. doi:10.1037/0096-3445.132.4.530

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, *52*, 463–477. doi:10.1016/j.jml.2004.12.001

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The university of South Florida free association, rhyme, and word fragment norms. Retrieved August 14, 2012, from http://w3.usf.edu/FreeAssociation/

Nelson, T. O. (1993). Judgments of learning and the allocation of study time. *Journal of Experimental Psychology: General*, *122*, 269–273. doi:10.1037/0096-3445.122.2.269

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. doi:10.1111/j.1467-9280.1994.tb00502.x

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.

Reder, L. M., & Anderson, J. R. (1982). Effects of spacing and embellishment on memory for the main points of a text. *Memory & Cognition*, *10*, 97–102. doi:10.3758/BF03209210

Roediger III, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255. doi:10.1111/j.1467-9280.2006.01693.x

Roediger III, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210. doi:10.1111/j.1745-6916.2006.00012.x

Rohrer, D. (2009). Avoidance of overlearning characterises the spacing effect. *European Journal of Cognitive Psychology*, *21*, 1001–1012. doi:10.1080/09541440802425954

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. In P. E. Morris & R. N. Sykes (Eds.), *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 907–912. doi:10.1037/0278-7393.27.4.907

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. doi:10.1037/0278-7393.30.3.601

Son, L. K., & Kornell, N. (2008). Research on the allocation of study time: Key studies from 1890 to the present (and beyond). In J. Dunlosky & R. A. Bjork (Eds.), *A handbook of memory and metamemory* (pp. 333–351). Hillsdale, NJ: Psychology Press.

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. doi:10.1037/0278-7393.26.1.204

Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin & Review*, *6*, 662–667. doi:10.3758/BF03212976

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. doi:10.1037/0278-7393.25.4.1024

Tullis, J. G., & Benjamin, A. S. (2012). Consequences of restudy choices in younger and older learners. *Psychonomic Bulletin & Review*, *19*, 743–749. doi:10.3758/s13423-012-0266-2

Valverde, G. A., & Schmidt, W. H. (1998). Refocusing US math and science education. *Issues in Science and Technology*, *14*, 60–66.

Vaughn, K. E., & Rawson, K. A. (2011). Diagnosing criterion-level effects on memory: What aspects of memory are enhanced by repeated retrieval? *Psychological Science*, *22*, 1127–1131. doi:10.1177/0956797611417724

Vaughn, K. E., & Rawson, K. A. (2014). Effects of criterion level on associative memory: Evidence for associative asymmetry. *Journal of Memory and Language*, *75*, 14–26. doi:10.1016/j.jml.2014.04.004

APPENDIX

Mean Accuracy on the Final Test Computed by Weighting All Items in a Given Condition Equally (Standard Deviation in Parentheses).

	Restudy known	*Restudy unknown*
Experiment 1		
Immediate	.687 (.176)	.906 (.075)
Delayed	.369 (.164)	.366 (.099)
Experiment 2		
Immediate	.574 (.218)	.748 (.171)
Delayed	.198 (.149)	.177 (.124)
Experiment 3		
Immediate	.566 (.287)	.773 (.233)
Delayed	.145 (.104)	.140 (.126)