Spaced Retrieval: Absolute Spacing Enhances Learning Regardless of Relative Spacing

Article *in* Journal of Experimental Psychology Learning Memory and Cognition · May 2011

DOI: 10.1037/a0023436 · Source: PubMed

2 authors:

Jeffrey D. Karpicke
Purdue University

Althea Bauernschmidt
St. Bonaventure University

Journal of Experimental Psychology:
Learning, Memory, and Cognition
2011, Vol. 37, No. 5, 1250–1257

0278-7393/11/$12.00 DOI: 10.1037/a0023436

RESEARCH REPORT

Spaced Retrieval: Absolute Spacing Enhances Learning Regardless of Relative Spacing

Jeffrey D. Karpicke and Althea Bauernschmidt
Purdue University

Repeated retrieval enhances long-term retention, and spaced repetition also enhances retention. A question with practical and theoretical significance is whether there are particular schedules of spaced retrieval (e.g., gradually expanding the interval between tests) that produce the best learning. In the present experiment, subjects studied and were tested on items until they could recall each one. They then practiced recalling the items on 3 repeated tests that were distributed according to one of several spacing schedules. Increasing the absolute (total) spacing of repeated tests produced large effects on long-term retention: Repeated retrieval with long intervals between each test produced a 200% improvement in long-term retention relative to repeated retrieval with no spacing between tests. However, there was no evidence that a particular relative spacing schedule (expanding, equal, or contracting) was inherently superior to another. Although expanding schedules afforded a pattern of increasing retrieval difficulty across repeated tests, this did not translate into gains in long-term retention. Repeated spaced retrieval had powerful effects on retention, but the relative schedule of repeated tests had no discernible impact.

Keywords: retrieval practice, testing effect, spacing effect, learning

Techniques that use spaced retrieval practice are becoming popular in many contexts. This is because the act of repeatedly retrieving knowledge produces powerful effects on learning (Roediger & Karpicke, 2006a, 2006b) and because spacing one's practice has long been known to enhance learning (see Greene, 2008).

Given the growing popularity of spaced retrieval techniques, it is important to determine whether there are particular schedules of retrieval practice that produce the best long-term retention. This topic is the focus of current controversy. In an important paper, Landauer and Bjork (1978) proposed that expanding retrieval schedules would produce better long-term retention than other schedules. In expanding schedules, people retrieve items immediately after studying them, thereby increasing the likelihood of successful recall, and then they gradually increase the spacing intervals between repeated tests. The rationale for expanding the intervals between repeated tests rests on two important assumptions. First, gradually expanding the interval between tests should increase the difficulty of repeated retrievals. Second, a pattern of increasing retrieval difficulty should enhance long-term retention (Gardiner, Craik, & Bleasdale, 1973; Pyc & Rawson, 2009).

The notion that expanding retrieval schedules would produce the best long-term retention gained acceptance after it was first proposed (for reviews, see Balota, Duchek, & Logan, 2007; Roediger & Karpicke, 2010). However, the results of several experiments have been mixed. Some experiments have obtained positive effects of expanding retrieval schedules relative to appropriate equally spaced control conditions (e.g., Cull, Shaughnessy, & Zechmeister, 1996, Experiments 1 and 4; Storm, Bjork, & Storm, 2010, Experiments 2 and 3). Others have found no differences between expanding and equally spaced conditions (e.g., Carpenter & DeLosh, 2005; Cull, 2000; Cull et al., 1996, Experiments 2, 3, and 5; Karpicke & Roediger, 2010; Logan & Balota, 2008; Pyc & Rawson, 2007; Storm et al., 2010, Experiments 1 and 3). Still other experiments have obtained advantages of equally spaced retrieval over expanding retrieval schedules (e.g., Cull, 2000; Karpicke & Roediger, 2007a; Logan & Balota, 2008).

The present research was carried out with three objectives in mind. First, we examined the effects of different relative spacing schedules (expanding, equally spaced, and contracting) across different levels of absolute or total spacing, as described below. Second, we used a new method to control and equate the level of retrieval success on initial tests across spacing conditions. Finally, we examined the theory that expanding retrieval schedules would produce patterns of retrieval difficulty that enhance long-term retention.

At the outset, it is important to make a distinction between the absolute and relative spacing of repeated tests. The absolute spacing of repetitions refers to the total number of trials that occur between all repeated tests, whereas relative spacing refers to how

This article was published Online First May 16, 2011.

Jeffrey D. Karpicke and Althea Bauernschmidt, Department of Psychological Sciences, Purdue University.

This research was supported in part by National Science Foundation Grant 0941170. We thank Jay Barot-Spencer and Samantha Ostler for their help collecting the data. We also thank James LeBreton and James Nairne for several helpful comments on this work.

Correspondence concerning this article should be addressed to Jeffrey D. Karpicke, Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2081. E-mail: karpicke@purdue.edu

the repeated tests are spaced relative to one another. For example, in an expanding condition a person might study an item and then try to recall it after one trial, then again after five more trials and again after nine more trials. This would be called a 1-5-9 schedule to reflect the number of trials that occurred between each test, and the absolute spacing of the repeated tests would be 15 trials. The appropriate comparison to this expanding condition is a condition matched on absolute spacing but differing in the relative spacing of tests. An equally spaced schedule where the interval between each test is five trials (a 5-5-5 schedule) has the same absolute spacing as the expanding condition but differs in the relative distribution of tests.

It is well known that increasing the absolute spacing between repetitions improves retention, especially in the long term (for review, see Greene, 2008). But the question of whether expanding retrieval is an inherently superior schedule is a question about the relative spacing of repetitions. Questions about absolute and relative spacing have traditionally been addressed in different literatures. Research on absolute spacing often examines the spacing of two study events, whereas research on relative spacing typically holds absolute spacing constant and examines the distribution of repeated tests (for review, see Roediger & Karpicke, 2010). In the present experiment, we examined expanding, equal, and contracting conditions that differed across levels of absolute spacing.

Another purpose in this experiment was to address a methodological problem that may cloud the inferences that can be drawn from spaced retrieval research. One feature of expanding retrieval schedules is that a first test occurs early to maximize the likelihood of retrieval success (and help ensure retrieval success on repeated tests). An interpretive problem often occurs in this comparison because subjects are bound to recall more items in expanding conditions with early first tests (e.g., after one trial in a 1-5-9 condition) than they would in equally spaced conditions with delayed first tests (e.g., after five trials in a 5-5-5 condition). For instance, subjects might successfully recall 80% of items on an immediate first test in an expanding condition and 60% of items on a delayed first test in an equally spaced condition (those means are estimated from Experiment 2 of Landauer and Bjork, 1978). In most spaced retrieval research, repeated tests occur without any feedback or opportunities to restudy following errors, because the purpose of the experiments is to examine direct effects of repeated retrieval, not mediated effects (Roediger & Karpicke, 2006a). Therefore, the position of the first test determines the probability of recall on that test and on all repeated tests. The consequence is that there are different levels of retrieval success across expanding and equally spaced conditions.

The presence of different levels of retrieval success across conditions clouds the interpretation of any effects that might occur on a later criterial test. If subjects initially recall more items in an expanding condition than they do in an equally spaced condition and if there is a difference between conditions on a final test, it is not clear to what source that difference should be attributed: the independent variable (spacing condition) or the different levels of retrieval success. Imagine a conceptually similar experiment in which one group of subjects restudies 80% of a list of words and is told to form mental images of the words, and another group restudies 60% of the words and is given rote repetition instructions. The interpretive problem is obvious: If the first group recalls more items than the second group, it is not clear whether the effect is due to the imagery instruction or to the fact that one group repeated more items. The same problem exists in spaced retrieval research. When there are different levels of retrieval success across relative spacing conditions, any effects may simply originate from that difference rather than from the spacing schedules.

Our solution to this problem was to have subjects study and test on items until they could successfully recall each one. Once an item was recalled, it was repeatedly tested three times according to one of several spacing schedules. Students learned foreign language vocabulary words across several study and test cycles. Once they recalled a word for the first time, it was assigned to a relative spacing condition and tested three additional times. This method of manipulating the spacing schedule of individual items minimizes the likelihood of different levels of retrieval success across spacing conditions (Karpicke, 2009; Karpicke & Roediger, 2008). The procedure also accurately represents what students do when they learn on their own: They continue studying and testing on items until they can recall them, rather than failing to recall some portion of items and not bothering to restudy them.

Many prior experiments have used a restricted range of spacing schedules similar to ones originally used by Landauer and Bjork (1978). In the present experiment, we varied both the absolute and relative spacing of repeated tests across several conditions, as shown in Table 1. The absolute spacing of repetitions was short (15 trials, which reflects conditions used in most prior research), medium (30 trials), or long (90 trials). Within each absolute spacing condition, individual items were assigned to one of three relative spacing conditions: an expanding schedule, an equally spaced schedule, or a contracting schedule in which the interval between repeated tests grew progressively shorter (Landauer & Bjork, 1978). As a control condition, we also included a "no spacing" group in which some items were studied only one time, some were studied and recalled to the criterion of one correct recall and dropped from further practice, and some were repeatedly recalled three times but in a massed fashion, with no spacing between repeated tests.

In addition, the majority of experiments on spaced retrieval have examined short-term retention, and only a few have examined retention after a delay of at least 1 day after learning (for reviews, see Balota et al., 2007; Roediger & Karpicke, 2010). In the present experiment, we examined the effects of these spaced retrieval conditions on a final 1-week delayed test.

As noted earlier, the theoretical rationale for expanding retrieval rests on two important assumptions: first, that expanding schedules

Table 1
Spacing Conditions Used in the Present Experiment

	Relative spacing		
Absolute spacing	Expanding	Equal	Contracting
Short (15)	1-5-9	5-5-5	9-5-1
Medium (30)	5-10-15	10-10-10	15-10-5
Long (90)	15-30-45	30-30-30	45-30-15

Note. Absolute spacing conditions were manipulated between subjects, and relative spacing conditions were manipulated within subjects. Numbers refer to the number of trials that occurred between each repeated test in the initial learning phase. There was also a no spacing control condition, described in the text.

will produce patterns of increasing retrieval difficulty across tests and, second, that patterns of increasing retrieval difficulty will be associated with greater levels of final recall. Although several studies have compared expanding and equally spaced retrieval schedules, few have examined these core theoretical ideas. To examine this theory directly, we recorded response times as measures of retrieval difficulty during initial tests (see Benjamin, Bjork, & Schwartz, 1998; Pyc & Rawson, 2009). We then examined (a) the relationship between spacing schedules and patterns of response times and (b) the relationship between patterns of response times and final recall.

The prediction, based on a wealth of research on the spacing effect, was that increasing the absolute spacing of tests would enhance retention. But given the mixed results in the literature on schedules of retrieval, whether relative schedules of repeated tests would matter for long-term retention was an open question. By controlling and equating the levels of retrieval success across relative spacing conditions, we might eliminate any effects of relative spacing on retention. Alternatively, if different relative spacing schedules afford different patterns of retrieval difficulty and if patterns of retrieval difficulty have consequences for long-term retention, there may be important effects of relative spacing schedules even once levels of retrieval success are equated across conditions.

Method

Subjects

Ninety-six Purdue University undergraduates participated in exchange for course credit.

Materials

The set of 100 Swahili–English word pairs in the Nelson and Dunlosky (1994) norms was used. Twenty-four pairs were used as critical to-be-learned pairs. The other 76 pairs were used as a pool of potential filler items during the learning phase.

Design

The experiment included a total of 12 spacing conditions. Absolute spacing was manipulated between subjects. There were four groups: short spacing, medium spacing, long spacing, and a no spacing control condition. Twenty-four subjects were assigned to each group. The absolute spacing of the three repeated tests was 15 trials, 30 trials, and 90 trials in the short, medium, and long conditions, respectively.

The relative spacing of repeated tests was manipulated within subjects. Once a word was recalled for the first time it was assigned to an expanding, equally spaced, or contracting schedule. The order in which items were assigned to relative spacing conditions was counterbalanced across subjects. Factorially crossing the three absolute spacing conditions with the three relative spacing conditions created the nine conditions shown in Table 1.

The no spacing control condition differed from the other conditions. In that condition, eight items were studied once and not repeatedly studied or tested (the study condition), eight items were studied and recalled to the criterion of one correct recall and then dropped from further practice (the recall once condition), and eight items were studied and recalled to the criterion of one correct recall and then repeatedly tested three times with zero trials between each test (the massed condition).

Procedure

Subjects were tested in small groups in two sessions. At the beginning of the experiment the subjects were told they would study and recall word pairs across a series of study and test trials. During study trials, subjects were shown a Swahili word with its English translation below it (e.g., *malkia–queen*) and were told to study the pair so they could remember the English word when given the Swahili word. During test trials, subjects were shown a Swahili word with a cursor below it and were told to type the correct English word for each Swahili word (e.g., subjects would be shown *malkia* as a cue to recall *queen*). Each study or test trial lasted 8 s with a 500-ms intertrial interval, after which the computer program advanced to the next trial, regardless of whether the subject had entered a response on test trials.

The learning phase consisted of a series of study and test trials that were mixed together, as is done in continuous paired-associate tasks (Karpicke & Roediger, 2007a). Subjects first studied all 24 pairs and then went through the list in a series of test trials. Once a subject recalled a word for the first time, the word was no longer presented in subsequent study trials, but it was repeatedly tested three times according to one of the three relative spacing schedules. If a subject failed to recall a word and had not yet recalled that word for the first time, the word pair was presented again during the next cycle of study trials and was tested again during the next cycle of test trials. If a subject studied and was tested on an item three times and still had not recalled it for the first time, the item was dropped from all further practice and deemed never recalled. This was done to prevent the learning phase from lasting an extremely long time. Across all subjects, only 4% of items were deemed never recalled.

The relative spacing schedule of the three repeated tests was determined by the computer program. For example, if an item was assigned to a 1-5-9 expanding schedule, the first repeated test was scheduled to occur one trial after the first recall, the second repeated test was scheduled to occur five trials after the first repeated test, and the third repeated test was scheduled to occur nine trials after the second repeated test. If a scheduling conflict occurred—that is, if an item was supposed to occur in a trial position that was already occupied by another item—the computer program placed the item in the next available trial position. Therefore, to resolve scheduling conflicts we allowed the actual spacings to increase beyond the nominal spacings (see Pashler, Zarow, & Triplett, 2003). If there was an empty trial position (without a study or test trial of a critical pair), the computer program presented a filler pair in a study trial in that trial position. Filler items were sampled without replacement from the pool of 76 possible items.

The learning phase terminated once all repeated tests had occurred. The subjects were then dismissed and returned for the final test 1 week later. The final test included the 24 critical pairs, tested in a random order, and did not include filler items. On the final test, the subjects were shown each Swahili word with a cursor below it and were told to type the correct English word for each

Swahili word. Each final test trial lasted 15 s (with a 500-ms intertrial interval). After they completed the final test, the subjects were debriefed and thanked for their participation.

Results

Relative Spacing Schedules

Table 2 shows the average actual lags assigned by the computer program and the total number of items assigned to each condition. The numbers of items assigned to each condition were roughly equivalent, and the actual spacings closely matched the intended spacings in each condition.

The average numbers of filler items used were 9.3, 17.1, and 44.5 in the short, medium, and long spacing conditions, respectively. No filler items were required in the no spacing condition.

Learning Phase Recall

Figure 1 shows the cumulative proportion of words recalled at least one time as measured in each test cycle (Karpicke & Roediger, 2007b). Subjects had recalled almost all words (96%) for the first time by the third cycle through the list (the third initial recall attempt), so 4% of words were never recalled and were not assigned to a spacing condition. The data from the three spaced conditions were entered into a 3 (absolute spacing: short, medium, or long) × 3 (test period) analysis of variance (ANOVA). There was a main effect of test period, $F(2, 138) = 857.07$, $\eta_p^2 = .93$, but no main effect of absolute spacing and no interaction (Fs < 1). Thus there were no differences in rate of learning across the three spaced conditions. However, subjects in the no spacing condition learned the list at a slightly slower rate than did subjects in the other conditions. The difference was significant on the first, $F(1, 94) = 6.72$, $\eta_p^2 = .07$, second, $F(1, 94) = 7.92$, $\eta_p^2 = .08$, and third test cycles, $F(1, 94) = 4.17$, $\eta_p^2 = .04$.

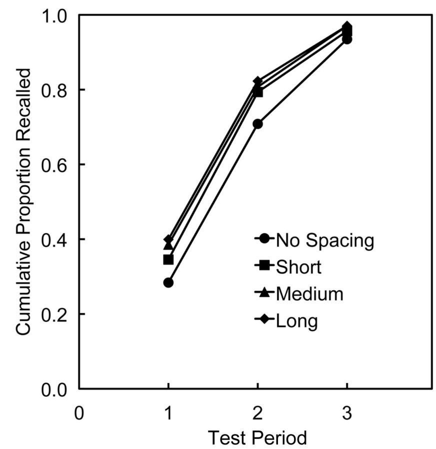

Figure 1. Cumulative proportion of words recalled as measured at each test cycle in the learning phase.

It was important to ensure that once subjects recalled items for the first time, they also successfully recalled the items on repeated tests. If subjects forgot items on the repeated tests, this would represent a failure to manipulate repeated retrieval practice. Table 3 shows the proportion recalled on repeated tests as a function of absolute and relative spacing conditions. The data from the three spaced conditions were collapsed across the three repeated tests and entered into a 3 (absolute spacing) × 3 (relative spacing) ANOVA. There was a main effect of absolute spacing, $F(2, 69) = 10.27$, $\eta_p^2 = .23$. Overall, the proportion of intertest retention in the long condition (.90) was lower than it was in the short condition (.97), $F(1, 46) = 13.54$, $\eta_p^2 = .23$, and medium condition (.96), $F(1, 46) = 9.51$, $\eta_p^2 = .17$, which did not themselves differ ($F < 1$). There was no main effect of relative spacing, $F(2, 138) = 1.28$, $p = .28$, and no interaction, $F(4, 138) = 1.78$, $p = .14$, which

Table 2
Total Number of Items Assigned to Each Relative Spacing Condition (Expanding, Equal, and Contracting) and the Mean Actual Lags Assigned by the Computer Program

Condition	No. items	Mean lag assigned by the computer program		
		Rep 1	Rep 2	Rep 3
Short spacing				
Expanding (1-5-9)	184	2.0	5.6	9.1
Equal (5-5-5)	184	5.6	5.5	5.0
Contracting (9-5-1)	183	9.6	5.0	1.2
Medium spacing				
Expanding (5-10-15)	187	5.7	10.4	15.1
Equal (10-10-10)	185	10.4	10.4	10.1
Contracting (15-10-5)	187	15.6	10.3	5.1
Long spacing				
Expanding (15-30-45)	186	15.4	30.3	45.1
Equal (30-30-30)	188	30.4	30.3	30.1
Contracting (45-30-15)	185	45.5	30.2	15.1

Note. The maximum number of items that could be assigned to each condition is 192 (24 subjects × 8 items). Rep 1, Rep 2, and Rep 3 refer to the first, second, and third repeated tests, respectively.

Table 3
Proportion of Words Recalled on the Three Repeated Tests in the Initial Learning Phase

Condition	Proportion recalled on repeated tests		
	Rep 1	Rep 2	Rep 3
No spacing			
Massed (0-0-0)	.97	.98	.97
Short spacing			
Expanding (1-5-9)	.97	.98	.98
Equal (5-5-5)	.97	.97	.97
Contracting (9-5-1)	.97	.97	.98
Medium spacing			
Expanding (5-10-15)	.97	.98	.98
Equal (10-10-10)	.95	.95	.95
Contracting (15-10-5)	.96	.97	.97
Long spacing			
Expanding (15-30-45)	.92	.92	.92
Equal (30-30-30)	.90	.91	.93
Contracting (45-30-15)	.87	.87	.88

Note. Rep 1, Rep 2, and Rep 3 refer to the first, second, and third repeated tests, respectively.

confirms that retrieval success on the repeated tests was equated across the three relative spacing conditions.

These analyses of learning phase performance indicate, first, that subjects recalled the majority of items in the learning phase and, second, that levels of repeated recall were quite high and did not differ across relative spacing conditions. Thus, the problem of different levels of retrieval success across spacing conditions was largely eliminated in this experiment.

Final Recall

Figure 2 shows the key results of this experiment: the proportion of words recalled on the final test as a function of absolute and relative spacing. There are several important results depicted in this figure. First, studying words only one time produced very poor final recall (1%), and studying and recalling words to the criterion of one correct recall in the recall once and massed conditions enhanced long-term retention. Repeatedly recalling items three times in a row in the massed condition produced no advantage beyond recalling items once (26% vs. 25%; $F < 1$). However, introducing spacing between the three repeated tests enhanced retention. All spaced retrieval conditions produced greater retention than the massed condition, Fs$(1, 46) > 13$, all $\eta_p^2 > .22$.

The critical questions were whether the relative spacing of repeated tests would affect long-term retention and whether any effects would depend on the absolute spacing of tests. There were clear effects of absolute spacing but no discernible effects of relative spacing on long-term retention. A 3 (absolute spacing) × 3 (relative spacing) ANOVA revealed a main effect of absolute spacing, $F(2, 69) = 11.07$, $\eta_p^2 = .24$. When we collapsed across relative spacing conditions, medium spacing produced greater recall than short spacing (64% vs. 49%), $F(1, 46) = 5.77$, $\eta_p^2 = .11$, and long spacing produced better recall than both the medium and short conditions (75% vs. 64%), $F(1, 46) = 4.98$, $\eta_p^2 = .10$; (75% vs. 49%), $F(1, 46) = 23.32$, $\eta_p^2 = .34$. Increasing the absolute spacing of repeated retrievals enhanced long-term retention.

However, there was no significant effect of relative spacing ($F < 1$) and no Absolute Spacing × Relative Spacing interaction ($F < 1$). To be sure that there were no effects of relative spacing, we conducted additional one-way ANOVAs on the three relative spacing conditions within each absolute condition. These analyses within the short, medium, and long conditions failed to indicate any effects of relative spacing (Fs < 1). In sum, varying the relative spacing of repeated tests did not produce any discernible effects on long-term retention.

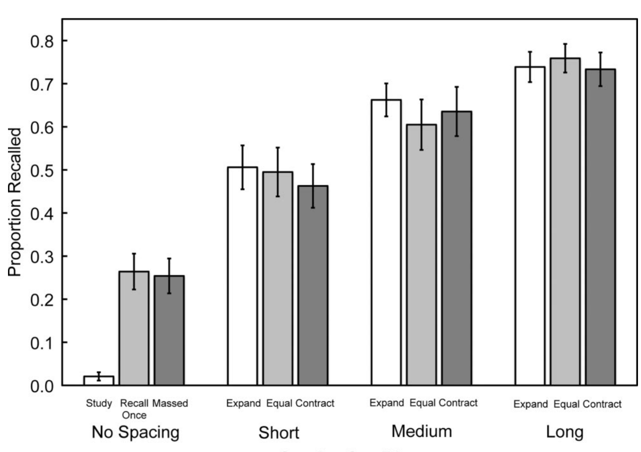

Figure 2. Proportion of words recalled on the final recall test 1 week after the learning phase. Bars indicate standard error.

The Relationship Between Spacing Conditions and Patterns of Response Times

In the last set of analyses, we turn to an examination of the theoretical rationale for expanding retrieval practice. The first component of the theory is that expanding schedules should afford patterns of increasing retrieval difficulty. If response times are considered measures of retrieval difficulty, a pattern of increasing response times would reflect increasing retrieval difficulty. The second component of the theory is that patterns of increasing retrieval difficulty should be associated with enhanced final recall. In the following analyses, only the short, medium, and long spacing conditions were considered. The no spacing conditions were not included.

Do different relative spacing conditions afford different patterns of retrieval difficulty across repeated tests? Response times were measured as the time between the onset of the cue and first keypress of the subject's response. Figure 3 shows the mean response times associated with each successful recall in the learning phase as a function of absolute and relative spacing condition. Once an item was recalled for the first time, response times decreased considerably on repeated tests, suggesting that repeated retrieval grew progressively easier even with very long lags between tests. An initial analysis indicated that response times for the first correct recalls did not differ across conditions ($F < 1$), which was to be expected because the spacing manipulation had not yet occurred. An analysis of mean response times is reported in the Appendix.

To examine how the pattern of response times across repeated tests varied as a function of relative spacing conditions, we calculated the slope of the best fitting least-squares regression line covering the response times on the three repeated tests. Table 4 shows the mean response time slopes, which were calculated for each item and then aggregated across subjects and conditions. Negative slopes would reflect patterns of increasing retrieval ease, whereas positive slopes (or relatively less-negative slopes) would reflect patterns of increasing retrieval difficulty.

A 3 (absolute condition) × 3 (relative condition) ANOVA on the response time slopes revealed a main effect of absolute spacing, $F(2, 69) = 18.83$, $\eta_p^2 = .35$, and a main effect of relative spacing, $F(2, 138) = 16.99$, $\eta_p^2 = .20$. The interaction approached significance, $F(2, 138) = 2.53$, $\eta_p^2 = .07$, $p = .08$. Pairwise analyses carried out within each absolute spacing condition indicated that there were significant differences among the slopes of all three relative spacing conditions in the short condition, Fs$(1, 23) > 4.56$, all $\eta_p^2 > .17$, and in the medium condition, Fs$(1, 23) > 3.79$, all $\eta_p^2 > .14$. None of the pairwise comparisons was significant in the long condition (Fs < 1).

In sum, relative spacing influenced the patterns of response times in the short and medium conditions. Expanding schedules afforded patterns of increasing retrieval difficulty, whereas equal

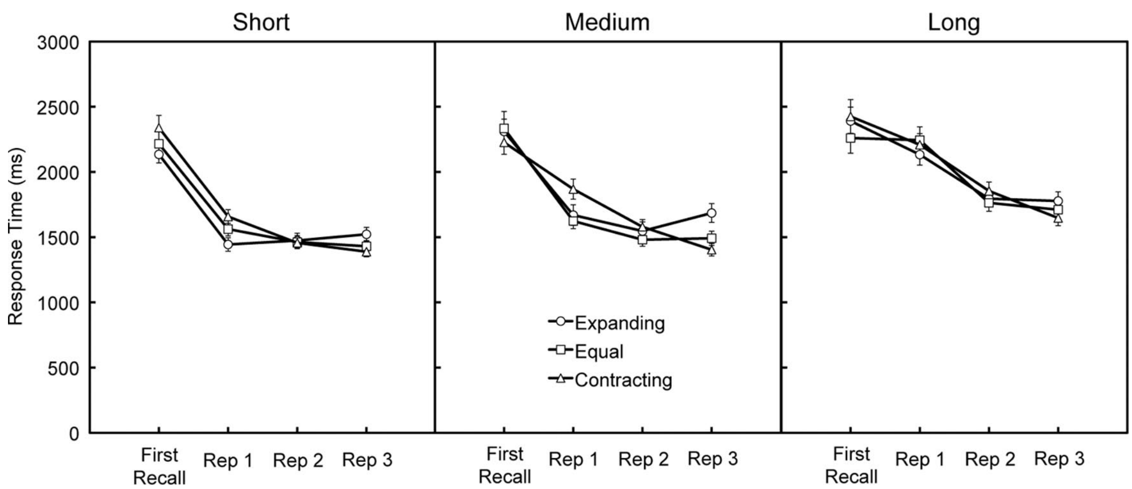

Figure 3. Average response times on the first recall and the three repeated recalls (Rep 1, Rep 2, and Rep 3) in the initial learning phase. Bars indicate standard error.

and contracting schedules produced patterns of decreasing difficulty. This result confirms that retrieval difficulty was successfully manipulated. However, variations in retrieval difficulty are important only insofar as they enhance retention, which we examine in the next section.

The Relationship Between Patterns of Response Times and Final Recall

To examine the relationship between patterns of response times and final recall, we conducted an analysis of covariance with absolute spacing condition, relative spacing condition, and response time slope as predictors and final recall as the dependent measure.[1] This analysis is similar to the ANOVA reported earlier with absolute and relative spacing conditions as predictors of final recall, but the analysis of covariance also assesses the influence of retrieval difficulty by including response time slope as a covariate. There was a main effect of absolute spacing, $F(2, 102.2) = 11.36$, $p < .01$, but no main effect of relative condition and no Absolute Spacing × Relative Spacing interaction (Fs < 1), replicating the earlier analysis. The main effect of response time slope approached but did not reach significance, $F(1, 169.1) = 2.61$, $p = .11$. None of the other interactions was significant (Fs < 1).

This analysis suggests that there may be a modest relationship between retrieval difficulty and retention. However, the relationship was in the opposite direction of what the expanding retrieval theory would predict. The Pearson correlation between slope and final recall, across subjects, was negative ($r = -.20$), indicating that negative slopes were associated with greater final recall. We also calculated the within-subject Pearson correlation between slope and recall to examine the difficulty-recall relationship for each subject at the individual-item level.[2] The average correlation was $r = .04$, which was not significantly different from zero, $t(70) = 1.35$, $p = .18$. Overall, the present analyses suggest that patterns of response times did not account for a large portion of the variance in final recall. If anything, patterns of increasing retrieval ease, not difficulty, tended to be associated with greater levels of final recall.

Table 4
Mean Slopes of the Best Fitting Least-Squares Regression Lines for Response Times on the Three Repeated Tests

Condition	Slope
Short spacing	
Expanding (1-5-9)	67.6
Equal (5-5-5)	−55.8
Contracting (9-5-1)	−146.8
Medium spacing	
Expanding (5-10-15)	23.4
Equal (10-10-10)	−77.7
Contracting (15-10-5)	−248.2
Long spacing	
Expanding (15-30-45)	−213.1
Equal (30-30-30)	−274.1
Contracting (45-30-15)	−266.9

Discussion

In this experiment, we examined several spaced retrieval schedules that varied in terms of both the absolute and relative spacing of repeated tests. The results confirmed that increasing the absolute spacing of repeated tests enhances long-term retention. Repeated retrieval with long intervals between each test produced a 200% improvement in final recall relative to repeated retrieval with three massed tests. However, whether repeated tests occurred in expanding, equally spaced, or contracting schedules did not produce any

[1] This analysis was carried out with the MIXED procedure in SPSS Version 17.0. Absolute spacing condition was a between-subjects factor, relative spacing condition was a within-subjects factor, and response time slope was a within-subjects covariate.

[2] The within-subject correlations were calculated for 71 subjects. One subject was excluded because he or she recalled all items on the final test. The correlations were calculated without regard to relative spacing conditions because our initial analysis showed that response time slopes did not interact with relative spacing conditions.

measurable impact on long-term retention. This finding contradicts the idea that expanding schedules should be inherently superior to other schedules, but the results are consistent with a growing literature that suggests that the relative spacing of tests may not matter much for long-term retention (Roediger & Karpicke, 2010).

Because of the way expanding and equally spaced conditions are usually implemented, expanding conditions often produce greater levels of initial retrieval success than do equally spaced conditions. This makes it unclear whether any effects on a criterial test are due to different spacing schedules or to different levels of retrieval success. By waiting until subjects had recalled items and then introducing the spacing manipulations, we were able to equate levels of retrieval success across conditions.

The original purpose of expanding retrieval was to balance the benefits of retrieval success and retrieval difficulty induced by spacing. One might worry that by bringing performance up to criterion, and thereby equating retrieval success across spacing conditions, we may have diluted the effects of retrieval difficulty on repeated tests. It is true that after items were recalled for the first time, retrieval grew considerably easier on repeated tests even with long lags between the tests. Only a few prior studies have examined whether different relative spacing schedules produce different patterns of retrieval difficulty, and all have found that retrieval grows easier across tests (Karpicke & Roediger, 2007a; Logan & Balota, 2008; Pyc & Rawson, 2009). However, when we examined the slope of response times across repeated tests, expanding schedules tended to produce patterns of increasing difficulty while equally spaced and contracting schedules did not. Therefore, even when retrieval success was equated across conditions, we still observed that expanding retrieval schedules tended to produce patterns of increasing retrieval difficulty.

However, these patterns of increasing retrieval difficulty were not associated with greater levels of final recall. Instead, patterns of increasing retrieval ease were associated with increases in final recall, although response time slopes did not account for a significant portion of the variance in final recall. Although some researchers have found that retrieval difficulty is associated with greater levels of final recall (Benjamin et al., 1998; Pyc & Rawson, 2009), others have found that initial retrieval ease is associated with greater recall (Koriat & Ma'ayan, 2005; Madigan, Neuse, & Roeber, 2000). In a similar vein, Jacoby, Shimizu, Daniels, and Rhodes (2005) showed that relatively easier decisions on an initial recognition memory test were associated with better performance on a subsequent recognition test. Most previous research has examined the relationship between the difficulty of a single initial retrieval and recall on a criterial test (see Benjamin et al., 1998; Gardiner et al., 1973). It is worth pointing out that in the present experiment, both the overall response times on initial tests and the probability of final recall increased as a function of absolute spacing (see the Appendix). But again, the rationale for expanding retrieval is not based on the overall retrieval difficulty afforded by the expanding condition; rather, it is based on the idea that a pattern of increasing retrieval difficulty across repeated tests should enhance final recall. In short, the relationship between patterns of retrieval difficulty across repeated tests and subsequent memory performance is not perfectly clear and certainly merits further scrutiny.

The present results confirm that spaced retrieval has robust effects on long-term retention. Spacing is a crucial component of repeated retrieval effects. Three repeated retrievals without any spacing (in the massed condition) was no better than recalling an item once, but introducing long spacing intervals between repeated tests produced a 200% gain in long-term retention. However, the particular way repeated tests were spaced relative to one another did not affect retention. This would appear to be good news for students, educators, and researchers interested in implementing spaced retrieval practice because it leads to a straightforward recommendation: Increasing the absolute spacing of retrieval attempts has clear value for learning, but how tests are spaced relative to one another may not be critical.

References

Balota, D. A., Duchek, J. M., & Logan, J. M. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extant literature. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger, III* (pp. 83–106). New York, NY: Psychology Press.

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. doi:10.1037/0096-3445.127.1.55

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology, 19,* 619–636. doi:10.1002/acp.1101

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14,* 215–235. doi:10.1002/(SICI)1099-0720(200005/06)14:3<215::AID-ACP640>3.0.CO;2-1

Cull, W. L., Shaughnessy, J. L., & Zechmeister, E. B. (1996). Expanding understanding of the expanding-pattern-of-retrieval mnemonic: Toward confidence in applicability. *Journal of Experimental Psychology: Applied, 2,* 365–378. doi:10.1037/1076-898X.2.4.365

Gardiner, J. M., Craik, F. I. M., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition, 1,* 213–216. doi:10.3758/BF03198098

Greene, R. L. (2008). Repetition and spacing effects. In H. L. Roediger (Ed.), *Learning and memory: A comprehensive reference. Vol. 2: Cognitive psychology of memory* (pp. 65–78). Oxford, England: Elsevier.

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review, 12,* 852–857. doi:10.3758/BF03196776

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, H. L. (2007a). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 704–719. doi:10.1037/0278-7393.33.4.704

Karpicke, J. D., & Roediger, H. L. (2007b). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57,* 151–162. doi:10.1016/j.jml.2006.09.004

Karpicke, J. D., & Roediger, H. L. (2008, February 15). The critical importance of retrieval for learning. *Science, 319,* 966–968. doi:10.1126/science.1152408

Karpicke, J. D., & Roediger, H. L. (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition, 38,* 116–124. doi:10.3758/MC.38.1.116

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language, 52,* 478–492. doi:10.1016/j.jml.2005.01.001

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London, England: Academic Press.

Logan, J. M., & Balota, D. A. (2008). Expanded vs. equal interval spaced retrieval practice: Exploring different schedules of spacing and retention interval in younger and older adults. *Aging, Neuropsychology, and Cognition, 15,* 257–280. doi:10.1080/13825580701322171

Madigan, S., Neuse, J., & Roeber, U. (2000). Retrieval latency and "at-risk" memories. *Memory & Cognition, 28,* 523–528. doi:10.3758/BF03201242

Nelson, T. O., & Dunlosky, J. (1994). Norms of paired-associate recall during multitrial learning learning of Swahili–English translation equivalents. *Memory, 2,* 325–335. doi:10.1080/09658219408258951

Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1051–1057. doi:10.1037/0278-7393.29.6.1051

Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition, 35,* 1917–1927. doi:10.3758/BF03192925

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60,* 437–447. doi:10.1016/j.jml.2009.01.004

Roediger, H. L., & Karpicke, J. D. (2006a). 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

Roediger, H. L., & Karpicke, J. D. (2006b). 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, H. L., & Karpicke, J. D. (2010). Intricacies of spaced retrieval: A resolution. In A. S. Benjamin (Ed.), *Successful remembering and successful forgetting: Essays in honor of Robert A. Bjork* (pp. 23–47). New York, NY: Psychology Press.

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition, 38,* 244–253. doi:10.3758/MC.38.2.244

Appendix

Analyses of Mean Response Times

The mean response times on the three repeated tests in the learning phase (labeled Rep 1, Rep 2, and Rep 3 in Figure 3) were entered into a 3 (absolute condition) × 3 (relative condition) × 3 (test number) ANOVA. There was a main effect of absolute condition, $F(2, 69) = 24.60$, $\eta_p^2 = .42$, indicating that average response times increased across the three absolute conditions (*M*s = 1,488, 1,594, and 1,903 ms in the short, medium, and long conditions, respectively). There was a main effect of test number, $F(2, 138) = 80.82$, $\eta_p^2 = .54$, indicating that response times tended to decrease across repeated tests (*M*s = 1,823, 1,601, and 1,562 ms on the first, second, and third repeated test, respectively). There was not a main effect of relative spacing or a significant Absolute Spacing × Relative Spacing interaction ($F\text{s} < 1$). However, test number interacted with absolute condition, $F(4, 138) = 14.58$, $\eta_p^2 = .30$, and relative condition, $F(4, 276) = 10.66$, $\eta_p^2 = .13$. Both interactions indicate that the rate at which response times changed across repeated tests depended on how the tests were spaced (cf. Karpicke & Roediger, 2007a). The three-way interaction was not significant, $F(8, 276) = 1.29$, $p = .25$.

This analysis confirms that the pattern of response times across repeated tests depended on the absolute and relative spacing conditions. The analyses of response time slopes presented in the text provide a direct examination of how patterns of response times differed across conditions.

Received August 3, 2010
Revision received February 4, 2011
Accepted March 1, 2011 ■