Psychological Bulletin
2006, Vol. 132, No. 3, 354–380

0033-2909/06/$12.00 DOI: 10.1037/0033-2909.132.3.354

Distributed Practice in Verbal Recall Tasks: A Review and Quantitative Synthesis

Nicholas J. Cepeda
University of California, San Diego and
University of Colorado at Boulder

Harold Pashler, Edward Vul, and John T. Wixted
University of California, San Diego

Doug Rohrer
University of South Florida

The authors performed a meta-analysis of the distributed practice effect to illuminate the effects of temporal variables that have been neglected in previous reviews. This review found 839 assessments of distributed practice in 317 experiments located in 184 articles. Effects of spacing (consecutive massed presentations vs. spaced learning episodes) and lag (less spaced vs. more spaced learning episodes) were examined, as were expanding interstudy interval (ISI) effects. Analyses suggest that ISI and retention interval operate jointly to affect final-test retention; specifically, the ISI producing maximal retention increased as retention interval increased. Areas needing future research and theoretical implications are discussed.

Keywords: spacing effect, distributed practice, meta-analysis, interstudy interval, retention interval

In the late 1800s, researchers began to demonstrate benefits from distributed practice (Ebbinghaus, 1885/1964; Jost, 1897; Thorndike, 1912). Since then, the topic of temporal distribution of practice has become one of the mainstays of learning and memory research. Recent reviews have suggested that a benefit from distributed practice is often found both for verbal memory tasks, such as list recall, paired associates, and paragraph recall (Janiszewski, Noel, & Sawyer, 2003), and for skill learning, such as mirror tracing or video game acquisition (Donovan & Radosevich, 1999). The size of the distributed practice effect is often large. In spite of abundant evidence for distributed practice benefits, a number of empirical studies (e.g., Toppino & Gracen, 1985; Underwood, 1961; Underwood & Ekstrand, 1967) and a recent review of the literature (Donovan & Radosevich, 1999) concluded that longer spacing and/or lag intervals sometimes failed to benefit retention. The present review explores the effects of distribution of practice upon retention of verbal information and seeks to elucidate the conditions under which distributed practice does and does not benefit retention.

Nicholas J. Cepeda, Department of Psychology, University of California, San Diego, and Department of Psychology, University of Colorado at Boulder; Harold Pashler, Edward Vul, and John T. Wixted, Department of Psychology, University of California, San Diego; Doug Rohrer, Department of Psychology, University of South Florida.

This work was supported by the Institute of Education Sciences (U.S. Department of Education Grants R305H020061 and R305H040108). We thank Jean Trinh for obtaining articles. We also thank Kelly Braun, Jane Childers, Michael Kahana, and Phil Pavlik for providing raw data. Finally, we thank Derek Briggs for comments on the article and statistical advice.

Correspondence concerning this article should be addressed to Nicholas J. Cepeda, University of Colorado at Boulder, Department of Psychology, 345 UCB, Boulder, CO 80309-0345. E-mail: ncepeda@psy.ucsd.edu

Terminology

The distributed practice effect refers to an effect of interstudy interval (ISI) upon learning, as measured on subsequent tests. ISI is the interval separating different study episodes of the same materials. In the most typical spacing study, there are two study episodes separated by an ISI and some retention interval separating the final study episode and a later test. Generally, the retention interval is fixed, and performance is compared for several different values of the ISI. In studies with more than two study episodes, retention interval still refers to the interval between the last of these study episodes and the final test.

When the study time devoted to any given item is not subject to any interruptions of intervening items or intervening time, learning is said to be *massed* (i.e., item A stays on the screen for twice as long as it would for a spaced presentation, without disappearing between presentations or disappearing for less than 1 s, such as the length of time it takes a slide projector to change slides). In contrast, learning is *spaced* or *distributed* when a measurable time lag (1 s or longer) separates study episodes for a given item—that is, either (a) item A appears, item A disappears for some amount of time, and then item A reappears or (b) item A appears, item A disappears, item B (item C, etc.) appears and disappears, and then item A reappears. For example, if a list of 20 items is presented twice, and there are no delays between each consecutive presentation of the list, learning episodes for any given item are spaced (on average) by 20 items, and this would be described as spaced learning. Learning is considered to be massed only when presentations of a given item in a list are separated by 0 items and a time lag of less than 1 s. During massed learning, the participant sees a single presentation of the item for twice the presentation time of a comparable spaced item. The term *spacing effect* refers to enhanced learning during spaced as compared with massed study episodes for a given item. In contrast, the term *lag effect* refers to

comparisons of different levels of spacing, either differing numbers of items (e.g., Thios & D'Agostino, 1976) or differing amounts of time (e.g., Tzeng, 1973). We use the generic term *distributed practice* to encompass both spacing and lag effects, without distinguishing between them.

As noted above, studies of distributed practice must include at least two, but may include more than two, learning episodes. When three or more learning episodes are presented, the ISIs may be equal (fixed), progressively longer (expanding), or progressively shorter (contracting).

Past Quantitative Reviews

The literature on distributed practice is vast, and the topic has been qualitatively reviewed in a number of books and articles (e.g., Crowder, 1976; Dempster, 1989; Greene, 1992; McGeoch & Irion, 1952; Ruch, 1928). Quantitative reviews are fewer in number: Four major quantitative reviews of distributed practice appear to exist (Donovan & Radosevich, 1999; Janiszewski, Noel, & Sawyer, 2003; T. D. Lee & Genovese, 1988; Moss, 1996). The authors of these articles all concluded that distributed practice produces an overall increase in retention, and they argued that the effect is moderated by several important variables. This section summarizes each of these reviews and highlights some of the questions that remain unanswered.

Moss (1996) reviewed 120 articles on the distributed practice effect, across a wide range of tasks. She partitioned data by age of participant and type of material (verbal information, intellectual skills, or motor learning). For each study, Moss determined the direction of effect, if any. She concluded that longer ISIs facilitate learning of verbal information (e.g., spelling) and motor skills (e.g., mirror tracing); in each case, over 80% of studies showed a distributed practice benefit. In contrast, only one third of intellectual skill (e.g., math computation) studies showed a benefit from distributed practice, and half showed no effect from distributed practice.

T. D. Lee and Genovese (1988) reviewed 47 articles on distributed practice in motor skill learning. Distributed practice improved both acquisition and retention of motor skills. (*Acquisition* refers to performance on the final learning trial, and *retention* refers to performance after a retention interval.) T. D. Lee and Genovese's findings dispute those of a prior review by Adams (1987; see also Doré & Hilgard, 1938; Irion, 1966). Adams's review concluded that distributed practice has little or no effect on acquisition of motor skills. In the 1960s, Hull's (1943) learning theory was shown to poorly account for existing data. Adams suggests that this discovery caused most researchers to stop studying the effects of distributed practice on motor learning. In contrast to Adams's claims and the 1960s negation of Hull' theory, both T. D. Lee and Genovese's (1988) review and Hull's theory suggested that distributed practice should improve motor learning.

In their meta-analysis of the distributed practice literature, Donovan and Radosevich (1999) inspected 63 articles that used a wide range of tasks. They examined the effects of several moderators: methodological rigor (on a 3-point scale), mental requirements (low or high, based on whether "mental or cognitive skills" [p. 798] were required for task performance), overall complexity (low, average, or high, based on the "number of distinct behaviors" [p. 798] required to perform the task), ISI (less than 1 min, 1–10 min, 10 min–1 hr, and greater than 1 day), and retention interval (less than or greater than 1 day). The largest effect sizes were seen in low rigor studies with low complexity tasks (e.g., rotary pursuit, typing, and peg reversal), and retention interval failed to influence effect size. The only interaction Donovan and Radosevich examined was the interaction of ISI and task domain. It is important to note that task domain moderated the distributed practice effect; depending on task domain and lag, an increase in ISI either increased or decreased effect size. Overall, Donovan and Radosevich found that increasingly distributed practice resulted in larger effect sizes for verbal tasks like free recall, foreign language, and verbal discrimination, but these tasks also showed an inverse-U function, such that very long lags produced smaller effect sizes. In contrast, increased lags produced smaller effect sizes for skill tasks like typing, gymnastics, and music performance. Thus, the current article is the first review article to suggest that distributed practice intervals can become too long, regardless of task domain. Their analysis omitted many articles that met their inclusion criteria (by our count, at least 55 articles that were published before 1999), and only about 10% of their sample used verbal memory tasks.

Janiszewski et al. (2003) performed the most extensive examination of distributed practice moderators to date; they focused on 97 articles from the verbal memory task literature. Five factors failed to influence effect size: verbal versus pictorial stimuli, novel versus familiar stimuli, unimodal versus bimodal stimulus presentation (e.g., auditory vs. auditory plus visual), structural versus semantic cue relationships, and isolated versus context-embedded stimuli. Five factors influenced effect size magnitude: lag (longer ISIs increased effect size), stimulus meaningfulness (meaningful stimuli showed a larger effect size than nonmeaningful stimuli), stimulus complexity (semantically complex stimuli showed a larger effect size than structurally complex or simple stimuli), learning type (intentional learning produced a larger effect size than incidental learning), and complexity of intervening material (intervening material that was semantically complex led to a larger effect size than intervening material that was structurally complex or simple). Unfortunately, Janiszewski et al. did not examine retention interval effects. Even though they focused on verbal memory tasks, there is only partial overlap between the articles used in Janiszewski et al.'s meta-analysis and those used in the present meta-analysis (47 articles were used in both). Partial overlap occurred in part because Janiszewski et al. chose to include studies that used reaction time, frequency judgments, and recognition memory as final-test learning measures, whereas we did not.

Summary of Past Quantitative Reviews

In summary, quantitative syntheses of the temporal distribution of practice literature have suggested that a benefit from longer ISIs is a fairly robust effect. Beyond that, however, few firm conclusions seem warranted. For example, Donovan and Radosevich's (1999) review suggested that increasingly distributed practice impairs learning, seemingly counter to Janiszewski et al.'s (2003) review, which concluded that increasingly distributed practice improved retention. Upon closer observation of Donovan and Radosevich's findings, skill acquisition studies showed decreased final-test learning with longer ISIs, and verbal memory tasks showed nonmonotonic effects of ISI on final-test learning (final-test performance improved as ISI increased from a few minutes to an hour and decreased as ISI reached 1 day or longer). Donovan and Radosevich's review suggested that retention interval has no

effect on the magnitude of the distributed practice effect. This conclusion is at variance with a number of individual experimental findings (e.g., Balota, Duchek, & Paullin, 1989; Bray, Robbins, & Witcher, 1976; Glenberg, 1976; Glenberg & Lehmann, 1980; see Crowder, 1976, for a useful discussion). Notably, Donovan and Radosevich failed to include in their meta-analysis many studies that showed retention interval effects. Even though distributed practice benefits are robust, temporal moderators affect distributed practice through a complex interplay of time and task.

Given the heterogeneity of studies included in prior syntheses, the omission of relevant studies, and the disparate conclusions of these syntheses, one might wonder whether they paint an accurate composite picture of the literature as a whole. In addition, prior syntheses have examined the joint impact of ISI and retention interval in a cursory fashion. If there is a complex interplay between ISI and retention interval, as some of the experimental studies cited in the previous paragraph would suggest, then this is likely to be of substantial import both for practical applications and for theoretical issues. The practical relevance is obvious: One can hardly select an ISI that optimizes instruction unless one knows how learning depends upon ISI; if that function varies with retention interval, this too must be considered in designing the most efficient procedures for pedagogy or training. Theories of the distributed practice effect are incomplete unless they can account for joint effects of ISI, retention interval, and task.

Learning and Relearning Confounds

One potentially critical factor that has been overlooked in past quantitative reviews of the distributed practice effect—potentially undermining many of the conclusions drawn—is the highly variable choice of training procedures used in the second and subsequent learning sessions. In many studies, including some deservedly well-cited research in this area (e.g., Bahrick, 1979; Bahrick & Phelps, 1987), participants were trained to a criterion of perfect performance on all items during the second and subsequent learning sessions. With this procedure, an increase in ISI inevitably increases the amount of training provided during the second or subsequent sessions. (This is because a longer ISI results in more forgetting between training sessions, thus necessitating a greater number of relearning trials to reach criterion.) Thus, in designs that have this feature, distribution of practice is confounded with the amount of practice time during the second (and subsequent) sessions. This makes it impossible to know whether differences in final-test performance reflect distributed practice effects per se. To avoid this confound, the number of relearning trials must be fixed. (Either training to a criterion of perfect performance during the first learning session or providing a fixed number of learning trials during the first learning session and then presenting items, with feedback, a fixed number of times during the second and subsequent learning sessions seems to us a reasonable way to equalize initial learning without introducing a relearning confound.)

Current Meta-Analysis

Our goal in the present article is to perform a quantitative integrative review of the distributed practice literature, tailored to shed light on the critical temporal and procedural variables discussed above. To examine ISI effects, we examined the degree of benefit produced by shorter and longer temporal gaps between learning episodes. We assessed joint effects of ISI and retention interval by examining ISI effects separately for a number of different retention intervals. Final-test performances following expanding versus fixed ISIs also were compared. In addition to providing additional clarity on the temporal variables just described, another goal of the present study was to pinpoint, for future research, important areas in which present distributed practice knowledge is severely limited. Although the literature on distributed practice is indeed very large, the present review discloses (in ways that previous reviews have not) how sorely lacking it is in the very sorts of information that are most needed if serious practical benefits are to be derived from this century-long research tradition.

We restricted our analysis to verbal memory tasks, in the broadest sense. These have been used in by far the greatest number of studies of distributed practice (Moss, 1996). This restriction was introduced because of the enormous heterogeneity of tasks and performance measures used in the remainder of the distributed practice literature. It seemed unlikely that the literature would allow meaningful synthetic conclusions to be drawn from any other single category of tasks or studies. Unlike previous reviewers, we restricted our review to studies using recall as a performance measure; we did not review studies that used performance measures like recognition or frequency judgments. To address potential relearning confounds, we examined the effects of providing different numbers of learning trials during the second session.

Method

Literature Search

Articles included in this analysis were selected by Nicholas J. Cepeda using several sources. Lists of potential articles were given to Nicholas J. Cepeda by Harold Pashler, Edward Vul, John T. Wixted, and Doug Rohrer, on the basis of past literature searches for related studies. PsycINFO (1872–2002) and/or ERIC (1966–2002) were searched with a variety of keywords. A partial list of keyword searches includes "spacing effect," "distributed practice," "spac* mass* practice," "spac* mass* learning," "spac* mass* presentation," "spac* mass* retention," "mass* distrib* retention," "spac* remem*," "distrib* remem*," "lag effect," "distrib* lag," "distrib* rehears*," "meta-analysis spacing," and "review spacing." Portions of article titles were entered as keywords into searches in these databases, and the resulting article lists were examined for potential articles. Primary authors were entered into PsycINFO searches, and their other articles were examined for relevance. Reference lists of all potential articles were examined for references to other potential studies. Reference lists from previous quantitative reviews (Donovan & Radosevich, 1999; Janiszewski et al., 2003; Moss, 1996) were examined. Internet searches were carried out (through http://www.google.com/) with the keywords "spacing effect" and "distributed practice." Current and older unpublished data were requested from researchers who (in our opinion) might be conducting distributed practice research or who might have older unpublished data.

Inclusion Criteria

Studies had to meet several criteria to be included. The material must have been learned during a verbal memory task (most commonly, paired-associates/cued recall, list recall, fact recall, or paragraph recall; also, text recall, object recall, sentence recall, spelling, face naming, picture naming, and category recall). A recall test must have assessed performance at the time of final test. The experiment must have provided two or more learning

opportunities for each item (or one learning opportunity of the same temporal length and separated by a lag less than 1 s, for massed items). Experiments using children and older adults were included (with some caveats noted below). Studies using clinical populations were excluded. Out of 427 reviewed articles, a total of 317 experiments in 184 articles met these criteria, providing 958 accuracy values, 839 assessments of distributed practice, and 169 effect sizes.

Data Coding

Time intervals were coded in days (e.g., 1 min = 0.000694 days, and 1 week = 7 days). ISI and retention interval were computed on the basis of authors' reports of either the number of items and/or the amount of time between learning episodes for a given item. When authors described lags in terms of the actual (or in some cases, typical) number of items intervening between learning episodes involving a given item, an estimate of the time interval was derived. If this estimate could not be derived, usually either because presentation time for items was not given or because there was too much variability in the number of items between learning episodes, the data were excluded. When an experimental procedure employed a list presentation, retention interval varied with serial position; thus, retention interval might be 10 s for one item and 1 min for another item. Because of this confound, we have reanalyzed the data, separating out list recall and paired associates studies (see the Appendix). For most analyses, data were separated into relatively small ranges of retention interval (e.g., less than 1 min, 1 min–less than 10 min, 10 min–less than 1 day, 1 day, 2–7 days, 8–30 days, 31 or more days. In some cases, the necessary temporal and/or accuracy data were not available in the published article, but we were able to obtain these data directly from the study author. For these studies, the reader will not be able to calculate ISI, retention interval, and/or accuracy from the published article.)

Computation of Effect Size

Cohen's *d* (Cohen, 1988) was selected as the measure of effect size, because of its widespread use in the literature. To calculate *d*, the difference in means was divided by the standard deviation.

Choice of standard deviation is crucial, as it impacts observed effect size (Glass, McGaw, & Smith, 1981; Taylor & White, 1992). Statisticians differ on the optimal type of standard deviation to use in computing effect size. Either control population standard deviation (Morris, 2000; Taylor & White, 1992) or various other forms of standard deviation (cf. D'Amico, Neilands, & Zambarano, 2001; Gleser & Olkin, 1994; Johnson & Eagly, 2000; Shadish & Haddock, 1994) are typically used. In this article, standard deviation was determined by use of the method advocated by D'Amico et al. (2001), whereby standard deviation at each ISI was calculated, and a simple average was taken across conditions in that experiment. Studies that failed to report enough information to calculate this form of standard deviation were excluded from effect size analyses.

In choosing to use this form of standard deviation, we implicitly assumed that experimental conditions had equal variance (Becker, 1988; Cohen, 1988). In reality, variance between conditions is rarely numerically equal. We feel that the present data adequately approximated this assumption, because rarely did variances at different ISIs differ by more than 10%. As well, most of the data examined here exhibit neither ceiling nor floor effects, a likely source of unequal variance.

For within-subject experiments, standard deviation was corrected for dependence between responses with the equation $SD_{ig} = SD_{ws}[2(1 - \rho)]^{1/2}$ from Morris and DeShon (2002; cf. Cortina & Nouri, 2000; Dunlap, Cortina, Vaslow, & Burke, 1996; Gibbons, Hedeker, & Davis, 1993), where SD_{ig} is the independent groups standard deviation, SD_{ws} is the within-subject standard deviation, and ρ is the correlation between scores. In the current analysis, correction for dependence used the average of all pairwise ISI correlations as input to the correction equation. When information necessary for this correction was unavailable, these data were excluded from effect size analyses.

Computation of ISI and Retention Interval Joint Effects

To examine the joint effects of ISI and retention interval, we performed three separate lag analyses. The first lag analysis was designed to mirror the lag analysis performed by Donovan and Radosevich (1999) and Janiszewski et al. (2003). This analysis does not allow claims about relative benefits of specific ISIs, for reasons that are described below. The second lag analysis does allow us to make claims about what specific ISI is optimal at each specific retention interval. The third (qualitative) lag analysis was designed to dispel concerns about a potential confound present in the first two lag analyses. In reading the following descriptions of absolute and difference lag analyses, the reader is referred to Figure 1.

Difference lag analyses. The first lag analysis was concerned with the differences in ISI and accuracy that are obtained when adjacent pairwise within-study experimental conditions are compared. For example, Figure 1 shows data from two hypothetical studies. Each study used ISIs of 1 min, 1 day, and 2 days. One study used a retention interval of 1 min, and the other study used a retention interval of 7 days. In performing difference lag analyses, we computed between-condition accuracy differences by subtracting the accuracy for the next shorter ISI from the accuracy value for the longer ISI: For each adjacent ISI pair from each study, accuracy difference = longer ISI accuracy − next short ISI accuracy. Likewise, the ISI difference was computed in the same way: For each adjacent ISI pair from each study, ISI difference = longer ISI − next shorter ISI.

Following the example in Figure 1, the ISIs used in Study 1 were 1 min, 1 day, and 2 days, resulting in two ISI differences. For ISIs of 2 days and 1 day, ISI difference = 2 days − 1 day = 1 day, and for ISIs of 1 day and 1 min, ISI difference = 1 day − 1 min = 1 day. Study 1 also yields two accuracy difference values. For ISIs of 2 days and 1 day, accuracy difference = 50 − 60 = −10%, and for ISIs of 1 day and 1 min, accuracy difference = 60 − 90 = −30%.

As seen in Figure 1, the average accuracy difference value for a retention interval of 1 min–2 hr and an ISI of 1 day is the mean of these two Study 1 accuracy difference values: −20%. The ISI difference and accuracy difference values for Study 2 are calculated and binned in a similar fashion.

ISI difference and accuracy difference values were calculated from all studies in the literature for which both difference values were calculable. When plotting each data point, we binned that data point with other data points using similar or identical ISI and retention interval values. For example, data points using an ISI of 2 days were averaged with data points using an ISI of 7 days (when their retention intervals were from the same bin as well).

We computed effect sizes by dividing each accuracy difference value by the appropriate standard deviation. After this uncorrected effect size was obtained, the corrections described in the *Computation of Effect Size* section were performed, when necessary. In many cases, standard deviation values were not available, and thus there are substantially fewer effect size data points than there are accuracy difference data points. (By grouping data into ISI bins in this manner, we lost the ability to draw conclusions about the relative benefits of specific ISIs. Instead, we were only able to make claims about the expected accuracy differences that would result if similar experimental manipulations of ISI had been used.)

Absolute lag analyses. Because we are interested in the relative benefits of specific ISIs, we also performed lag analyses on the basis of absolute accuracy at specific ISIs and retention intervals. To compute absolute lag effects, we first binned data into varying ranges of ISI and retention interval. We then averaged the accuracy values from every data point within each ISI and retention interval bin. Referring again to the hypothetical data in Figure 1, Study 1 used ISIs of 1 min, 1 day, and 2 days. One accuracy value (the accuracy at ISI = 1 day; 60% correct) would be placed into the ISI = 1 day, retention interval = 1 min–2 hr bin; another accuracy value (the accuracy at ISI = 2 days) would be placed into the ISI = 2–28 days, retention interval = 1 min–2 hr bin. Each study in Figure 1 yields three accuracy values that are grouped into ISI and retention interval bins. (Note that each study in Figure 1 yielded one accuracy difference value for the difference lag analyses.)

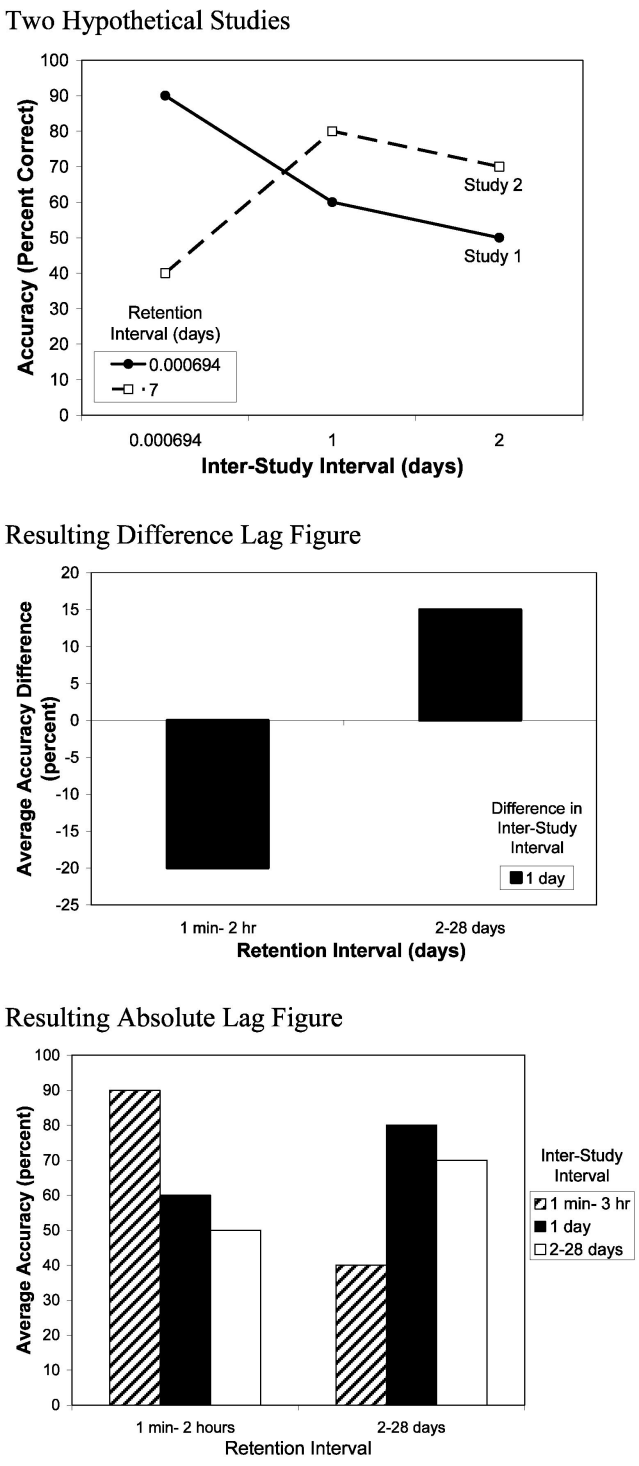

Figure 1. Graphical representation of two hypothetical studies and the difference and absolute lag graphs that would result when lag analysis of these studies is performed.

To determine the relative benefits of specific ISIs, we were interested in the changes in average accuracy across different ISI bins, for a given retention interval bin. However, different studies contribute data to each ISI bin, even within a given retention interval bin. Thus, our comparisons of interest, for both difference and absolute lag analyses, involved between-study comparisons. This was problematic, as overall level of difficulty often differed substantially between studies. Because we did not correct for these differences, the overall level of difficulty may not be equivalent for every bin. Thus, both absolute and difference analyses were confounded. This confound was present in prior meta-analyses as well. Because of our concerns about this confound, we performed an additional analysis, which uses within-study instead of between-studies methods to determine how optimal ISI changes with retention interval. This third analysis method does not include the just-described confound.

Within-study lag analyses. As a third method for determining if and how optimal ISI changes as a function of retention interval, we qualitatively examined studies that included an optimal ISI. Studies with an optimal ISI are those that included at least three different ISI conditions, wherein one ISI condition had an accuracy value higher than the immediately shorter ISI and which was immediately followed by a longer ISI condition with an equal or lower accuracy value. Thus, the optimal ISI can be described as the shortest ISI that produced maximal retention. We examined whether these optimal ISIs were longer for longer retention intervals. (This analysis is subject to some caveats: First, it may be that the highest accuracy in a study is a local maximum and that another ISI would have produced higher accuracy had more ISIs been used in the study. The smaller the range of absolute ISIs used, the greater is this potential problem. Second, the actual observed optimal ISI varies, as not all ISIs were tested within a given study. The degree to which the observed optimal ISI might vary from the truly optimal ISI depends on the distance between the immediately adjacent ISI values. Even with these caveats, we believe that this analysis provides a good estimate of optimal ISI.)

Results and Discussion

Analyses examined the joint effects of ISI and retention interval on final-test retention, as well as the effects of massed versus spaced learning. We examined joint effects of ISI and retention interval separately for paired associate and list recall tasks, and we examined qualitative differences between studies—specifically, the influence of experimental design, relearning method, and expanding study intervals.

Spacing Effects: Massing Versus Spacing

The spacing effect hinges upon a comparison of massed and spaced presentations of a to-be-learned item. (As noted above, if a list of items was presented twice in immediate succession, this was considered a spaced presentation, because the learning of any given item took place on two different occasions in time. To qualify as a massed presentation, there must have been either a single uninterrupted presentation of the item during learning or a lag shorter than 1 s.) Our analysis of massed versus spaced learning compared massed learning with the shortest spaced learning interval provided within a given study. Studies that failed to include a massed presentation were excluded, leaving 271 comparisons of retention accuracy and 23 effect sizes. Only accuracy differences are reported, because of insufficient effect size data. Independent samples t tests were used for analyses, as a conservative measure, as some studies were between subjects and others were within subject.

Spaced presentations led to markedly better final-test performance, compared with massed presentations. For retention intervals less than 1 min, spaced presentations improved final-test performance by 9%, compared with massed presentations (see Table 1). This finding appears to run counter to what has sometimes been referred to as the "Peterson paradox," wherein there is purportedly a massing benefit at short retention intervals. Perhaps

Table 1
Percentage Correct on the Final Recall Test for Massed and Spaced Conditions, Number of Performance Differences and Studies, Total Number of Participants Summing Across All Study/Condition Combinations, and Statistical Analyses, for Spaced Versus Massed Presentations

Retention interval	% Correct			No. of performance differences	No. of studies	No. of participants	Statistical analysis
	Massed	Spaced	*SE*				
1–59 s	41.2	50.1	1.7	105	96	5,086	$t(208) = 3.7$, $p < .001$
1 min–less than 10 min	33.8	44.8	1.5	124	117	6,762	$t(246) = 5.0$, $p < .001$
10 min–less than 1 day	40.6	47.9	8.1	11	10	870	$t(20) = 0.6$, $p = .535$
1 day	32.9	43.0	6.0	15	15	1,123	$t(28) = 1.2$, $p = .249$
2–7 days	31.1	45.4	7.3	9	9	435	$t(16) = 1.4$, $p = .190$
8–30 days	32.8	62.2	8.8	6	6	492	$t(10) = 2.3$, $p < .05$
31 days or more	17.0	39.0		1	1	43	
All retention intervals	36.7	47.3	1.1	271	254	14,811	$t(540) = 6.6$, $p < .001$

this massing benefit occurs only with extremely short retention intervals. For example, Peterson, Hillner, and Saltzman (1962) found a massing benefit only when retention interval was 2 or 4 s and not when retention interval was 8 or 16 s. Similarly, Peterson, Saltzman, Hillner, and Land (1962) found a massing benefit at retention intervals of 4 and 8 s, but Peterson, Wampler, Kirkpatrick, and Saltzman (1963) failed to find a massing benefit at retention intervals of 8, 16, or 60 s. All these studies used very short ISIs, from 4 to 8 s. (The two tasks most predominantly used by researchers—paired associate and list learning—were well represented across retention intervals.) Only 12 of 271 comparisons of massed and spaced performance showed no effect or a negative effect from spacing, making the spacing effect quite robust. Most of these 12 comparisons used the same task type as studies that did show a spacing benefit—paired associate learning.

We examined the interaction between magnitude of the spacing effect and retention interval by calculating the difference in performance between massed and spaced presentations and collapsing over each of seven retention interval ranges (see Table 1); there is no hint that massed presentation was preferable to spaced, whether retention interval was very short (less than 1 min) or very long (over 30 days). This suggests that there is always a large benefit when information is studied on two separate occasions instead of only once. (Note that in every case examined here, the amounts of study time for massed and spaced items were equivalent; thus, this spacing benefit was not due to presentation time.)

Lag Effects: Joint Effects of ISI and Retention Interval

Lag effects refer to changes in final-test memory performance as a function of change in ISI, when both ISIs and the differences between ISIs are greater than 0 s (in the current data set, at least 1 s). Prior reviews (Donovan & Radosevich, 1999; Janiszewski et al., 2003) found different relationships between ISI and effect size; Donovan and Radosevich (1999) reported nonmonotonic effects of ISI difference on effect size, whereas Janiszewski et al. (2003) found an increase in effect size as ISI difference increased. We have extended these previous reviews by including both ISI difference and retention interval in our analysis. It is possible that Donovan and Radosevich and Janiszewski et al. found these different patterns because the optimal ISI difference changes as a function of retention interval, and their reviews happened to include studies using different retention intervals. It is also possible that prior meta-analyses' use of ISI differences rather than absolute ISIs influenced their findings, as information is lost during difference computation. (Unfortunately, we do not have access to the actual data used in each review and thus cannot test these predictions directly.)

To examine how absolute ISI and ISI difference interacts with retention interval, we grouped the accuracy data into bins with boundaries varying roughly by one log order of magnitude (limited by the amount of data available). We would have preferred to use more precise log orders of magnitude to create our bins, but combinations of ISI difference and retention interval are not evenly represented by the existing literature. Figure 2 plots each ISI difference and retention interval combination from every study included in our difference lag analyses. If this combination space were evenly represented, Figure 2 would show a uniform "cloud" of data points. In addition to the irregular sampling of ISI difference and retention interval combinations, large subsets of this combination space contain sparse amounts of data, or are missing data altogether. To best utilize the full range of data, we created our own ISI and retention interval bins in a way that maximized data usage while still attempting to capture log order of magnitude changes.

Accuracy difference and effect size lag analyses. The vast majority of mean performance differences (80%) used a retention interval of less than 1 day, and only a few differences (4%) used a retention interval longer than 1 month (see Table 2). As mentioned earlier, Figure 2 shows this failure of the literature to fully represent the space of ISI and retention interval combinations. This feature of the literature impacts our ability to analyze the qualitative findings from our difference lag analyses with inferential statistics. (A recent case study critiquing meta-analysis technique suggests that statistical testing is not necessary to produce valid, interpretable findings; Briggs, 2005).

For each study, we computed the accuracy difference that resulted from each pairwise ISI difference, and we plotted the average of these accuracy differences as a function of ISI difference and retention interval (see Figure 3). Only ISI difference by retention interval bins that include three or more mean performance differences are shown. Several bins have fewer than three mean accuracy differences, and accuracy difference values from bins with at least one data point are qualitatively consistent with the pattern of results shown in Figure 3. There is little, if any, ISI

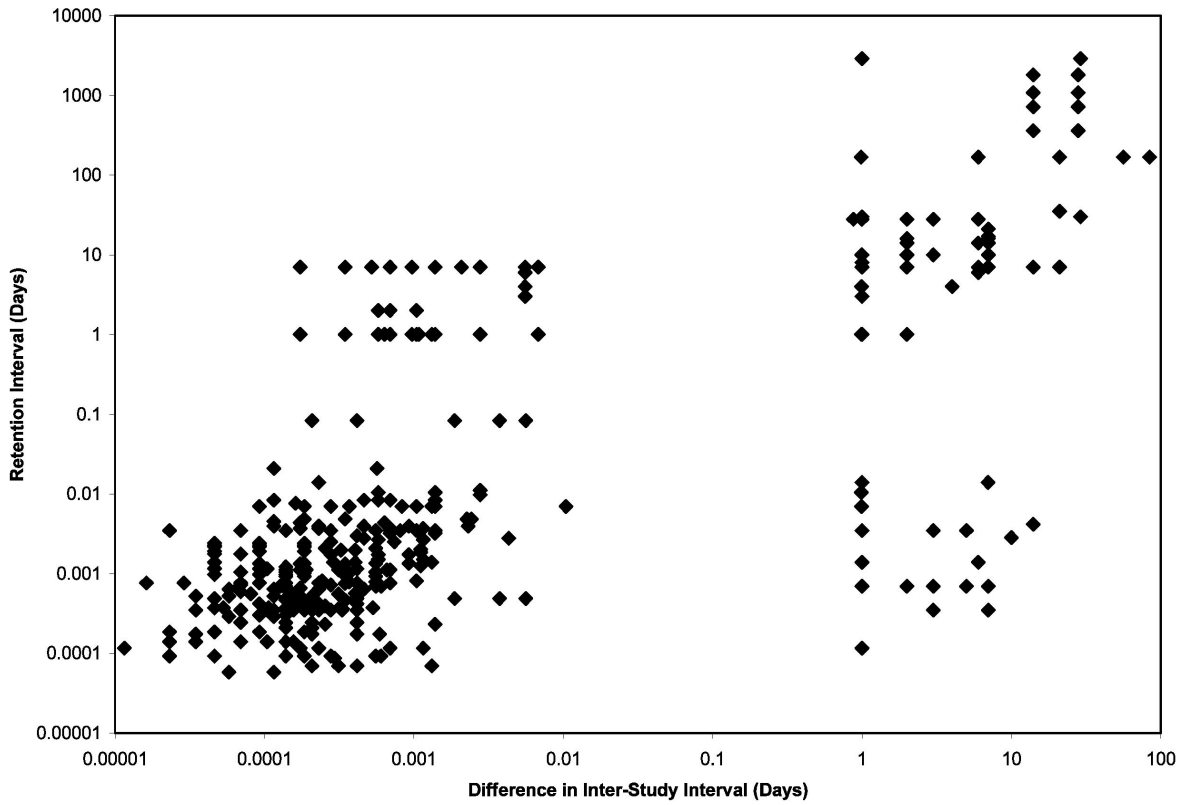

Figure 2. Scatter plot of interstudy interval (ISI) difference by retention interval, for all studies in the accuracy difference lag analyses.

difference effect at retention intervals shorter than 1 day. In sharp contrast, for a 1-day retention interval, performance significantly increased as ISI difference increased from 1–15 min to 1 day. Qualitatively, one study suggested that performance should drop when ISI difference increases beyond 1 day. The same pattern of results is seen with a 2- to 28-day retention interval: A 1-day ISI difference produced a significant benefit over the 1- to 15-day ISI, and there was a marginally significant drop in performance as ISI difference increased beyond 1 day. For retention intervals longer than 1 month, we must rely on qualitative results, which suggest that the optimal ISI difference is longer than 1 day at retention intervals longer than 1 month. Overall, the results show a tendency for the greatest increases in final-test recall to be found at longer ISI differences, the longer the retention interval. The qualitative pattern that optimal ISI difference increases as retention interval increases is supported by quantitative analyses of the bin data (see Table 3). Furthermore, effect size data mirror these findings from the accuracy data (see Figure 4).

Table 2
Number of Performance Differences, Data Points, and Effect Sizes, for Accuracy Difference, Absolute, and Effect Size Lag Analyses, Respectively, by Retention Interval Range

Retention interval range	No. of performance differences	No. of data points	No. of effect sizes
2–59 s	174	301	14
1 min–2 hr	259	452	53
1 day	27	52	16
2–28 days	56	108	31
30 days or more	23	34	19

Portions of our data are qualitatively similar to other meta-analysis findings. Like Donovan and Radosevich's (1999) data, our data show nonmonotonic effects of ISI difference. Like Janiszewski et al.'s (2003) results, our data show generally improved retention as ISI difference increases. Unfortunately, it is impossible to know whether we have confirmed these meta-analyses, because we do not know the retention interval values used in each prior meta-analysis; however, our results provide a plausible mechanism by which these prior discrepant findings might be reconciled.

For accuracy data, which are depicted in Figure 3, Table 4 shows the number of data points that use paired associate, list recall, or other types of tasks, and the overall number of data points, studies, and unique participants included in each bin. If the relative percentage of data points using each type of task changes between bins, then changes in optimal ISI difference with change in retention interval could potentially be due to changes in the percentage of data points using each task type as opposed to changes in retention interval. In the Appendix, Figures A1 and A2 (for paired associate and list recall tasks, respectively) illustrate that the joint effects of ISI difference and retention interval are due to changes in retention interval and not to changes in task type.

Absolute ISI lag analyses. Although it is encouraging that difference lag analyses show clear joint effects of ISI difference

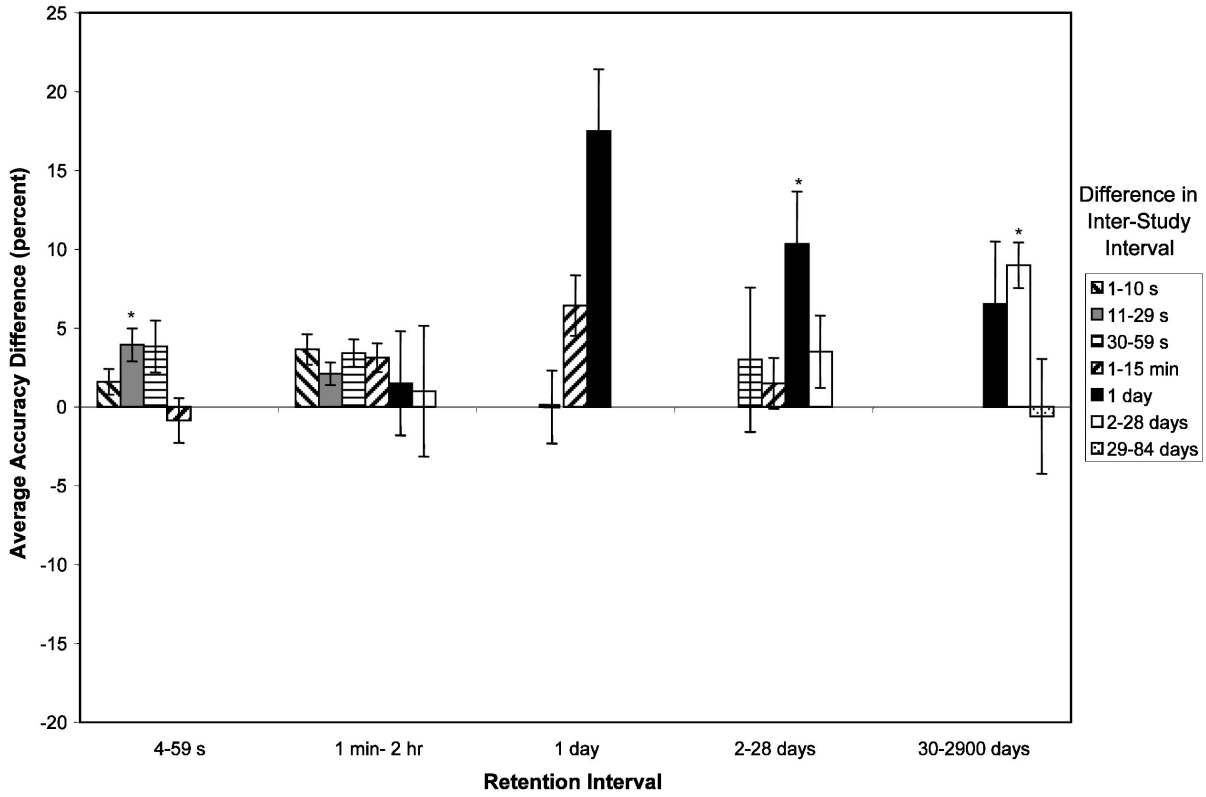

Figure 3. For all studies in the accuracy difference lag analyses, accuracy difference between all adjacent pairs of interstudy interval (ISI) values from each study, binned by difference in ISI and retention interval and averaged across studies. When surrounded by ISI bins with lower accuracy values, the ISI bin showing the highest accuracy value at each retention interval bin is indicated with an asterisk. Error bars represent one standard error of the mean.

and retention interval, we are really interested in how absolute ISI interacts with retention interval. On the basis of the absolute optimal ISI data, we can make concrete recommendations on how large of a lag is optimal, given a particular retention interval. Differences in performance between optimal and suboptimal ISI differences should be smaller and less meaningful as a measure of ideal absolute ISI, compared with differences between optimal and suboptimal absolute ISIs. This is the case because ISI differences of 7–8 days and ISI differences of 0–1 day are combined in difference ISI analyses but not in absolute lag analyses, and we would expect an ISI change from 0 to 1 day to show a much larger effect than an ISI change from 7 to 8 days.

Mirroring accuracy difference data, most data points used a retention interval less than 1 day, and only a few data points used a retention interval longer than 1 month (see Table 2). Just as the literature failed to represent the full combination space of ISI differences and retention intervals for the difference lag analyses, so too was the space of ISI and retention interval

Table 3
Shorter and Longer Interstudy Interval (ISI) Range, Retention Interval Range, Percentage Correct at the Shorter and Longer ISI Range, and Statistical Analyses, for Accuracy Difference Lag Analyses

Shorter ISI range	Longer ISI range	Retention interval range	% correct at ISI range			Statistical analysis
			Shorter	Longer	*SEM*	
1–10 s	11–29 s	4–59 s	1.6	3.9	0.9	$t(147) = 1.8$, $p = .077$
11–29 s	1–15 min	4–59 s	3.9	−0.9	1.2	$t(75) = 1.4$, $p = .156$
30–59 s	1 day	1 min–2 hr	3.4	1.0	2.5	$t(61) = 0.9$, $p = .397$
1–15 min	1 day	1 day	6.4	17.5	2.9	$t(16) = 2.7$, $p < .05$
1–15 min	1 day	2–28 days	1.5	10.3	2.5	$t(26) = 2.4$, $p < .05$
1 day	2–28 days	2–28 days	10.3	3.5	2.8	$t(37) = 1.7$, $p = .091$
1 day	2–28 days	30–2,900 days	6.5	9.0	2.7	$t(15) = 0.7$, $p = .476$
2–28 days	29–84 days	30–2,900 days	9.0	−0.6	2.6	$t(17) = 3.0$, $p < .01$

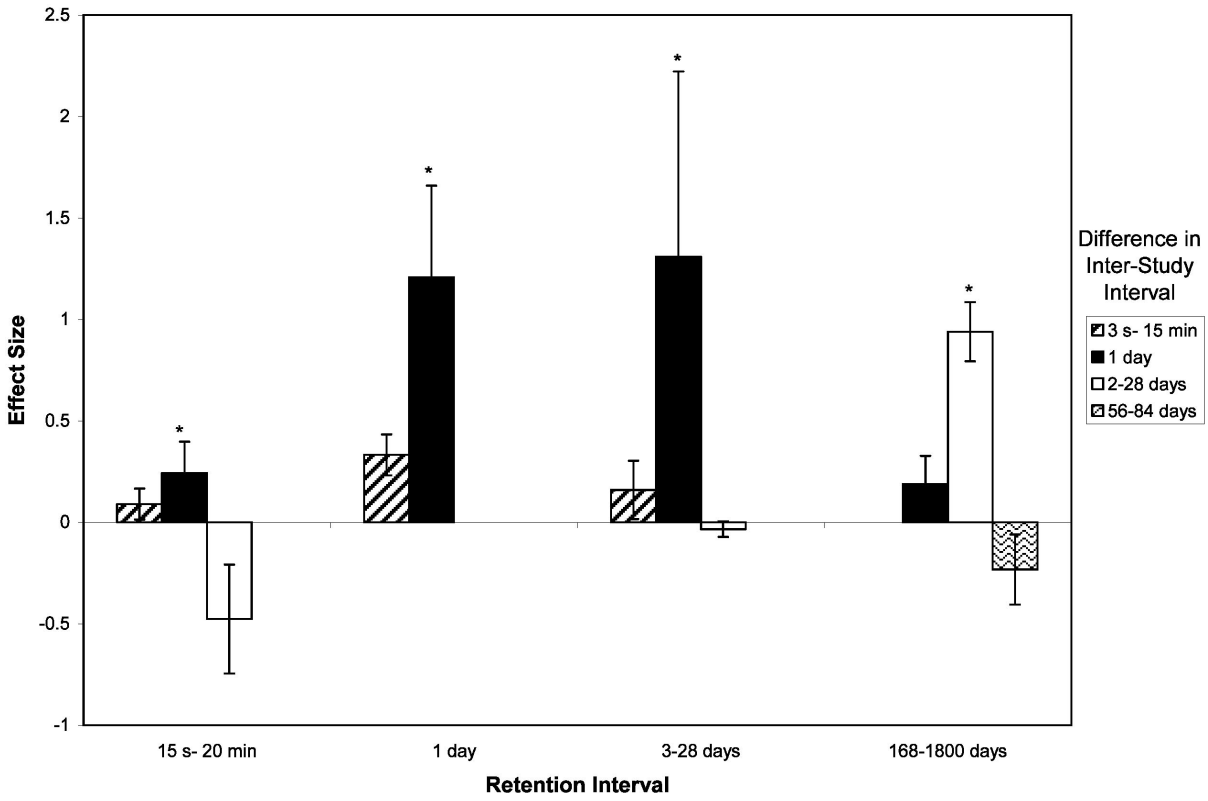

Figure 4. For all studies in the effect size lag analyses, effect sizes for all adjacent pairs of interstudy interval (ISI) values from each study, binned by difference in ISI and retention interval and averaged across studies. When surrounded by ISI bins with smaller effect size values, the ISI bin showing the largest effect size at each retention interval bin is indicated with an asterisk. Error bars represent one standard error of the mean.

combinations inadequately sampled for the absolute lag analyses (see Figure 5).

The plot of absolute ISI bin by retention interval bin is similar to the plot of ISI difference bin by retention interval bin (compare Figures 6 and 3). Although there are small differences in the ISI bin showing optimal performance, in both cases, the trend is for the optimal ISI bin to increase as retention interval increases. Quantitative analyses are shown in Table 5, and the number of data points that used each task type is shown in Table 6. In the Appendix, data are separated by task type, either paired associate or list recall. As in the ISI difference lag analysis, only absolute ISI by retention interval bins that include three or more data points are shown.

Within-study lag analyses. One problem with our absolute and difference lag analyses is that different studies contribute differentially to each bin. That is, each bin does not represent the same combination of studies. For this reason, one must be wary that task difficulty or other study-related factors played a role in differences between bins. A better comparison of lag effects would come from within-study comparisons, across a wide range of ISIs and retention intervals, as this eliminates the problem with task difficulty. To date, this massive study, which would need to include dozens of ISI and retention interval combinations, has not been conducted. Nonetheless, individual studies that represent a wide range of ISIs, both sub- and supraday, at a single retention interval, are supportive of our findings: Cepeda et al. (2005) presented data in which the optimal ISI was longer than 1 day at a supramonth retention interval; Gordon (1925) showed that subday ISIs are optimal at subday retention intervals and that supraday ISIs are optimal at supraday retention intervals; Glenberg and Lehmann (1980) showed results that mirror those of Gordon. These three studies are consistent with a number of other studies (e.g., Balota, Duchek, & Paullin, 1989; Glenberg, 1976; Peterson, Wampler, Kirkpatrick, & Saltzman, 1963) that show within-study support for the hypothesis that optimal ISI increases as retention interval increases. Table 7 shows results for individual studies that examined ISIs and retention intervals of 1 day or more.

Lag analysis summary. In summary, synthetic analyses support the robustness and generality of ISI and retention interval joint effects that a few oft-cited individual experiments have sometimes observed. Whereas earlier quantitative syntheses had sought to uncover effects of ISI difference or retention interval per se, the present review suggests that the literature as a whole reflects nonmonotonic effect of absolute ISI upon memory performance at a given retention interval, as well as the positive relationship between retention interval and the optimal absolute ISI value for that retention interval.

Experimental Design Issues

As noted in the introductory section, in examining commonly used experimental designs, we found that a number of frequently

Table 4
Number of Performance Differences and Studies, Unique Participants, and Performance Differences Using Paired Associate, List Recall, or Other Task Types, for Accuracy Difference Lag Analyses, by Retention Interval Range and Interstudy Interval (ISI) Range

Retention interval range	ISI range	No. of performance differences	No. of studies	No. of unique participants	No. using		
					Paired associate tasks	List recall tasks	Other tasks
4–59 s	1–10 s	79	28	1,539	35	41	3
4–59 s	11–29 s	70	39	2,083	20	48	2
4–59 s	30–59 s	18	12	694	6	12	0
4–59 s	1–15 min	7	4	327	5	2	0
1 min–2 hr	1–10 s	43	25	1,384	10	21	12
1 min–2 hr	11–29 s	91	50	2,736	27	59	5
1 min–2 hr	30–59 s	50	41	2,478	13	27	10
1 min–2 hr	1–15 min	52	40	3,295	18	13	21
1 min–2 hr	1 day	10	7	180	2	5	3
1 min–2 hr	2–28 days	13	9	618	3	0	10
1 day	30–59 s	9	8	469	4	5	0
1 day	1–15 min	14	9	667	6	6	2
1 day	1 day	4	4	86	0	3	1
2–28 days	30–59 s	3	3	174	1	0	2
2–28 days	1–15 min	14	6	613	6	6	2
2–28 days	1 day	14	11	902	5	4	5
2–28 days	2–28 days	25	18	4,118	7	6	12
30–2,900 days	1 day	4	3	106	4	0	0
30–2,900 days	2–28 days	13	3	294	12	0	1
30–2,900 days	29–84 days	6	3	160	6	0	0

cited studies contained serious design confounds or failed to implement the claimed experimental manipulation. Given their obvious practical importance, we specifically examined studies that used ISIs and retention intervals of 1 or more days (i.e., the studies in Table 7), to assess the quality of each study.

Studies contained several different confounds. One group of studies provided learning to perfect performance and then relearning, with feedback, to the criteria of perfect performance (Bahrick, 1979; Bahrick et al., 1993; Bahrick & Phelps, 1987). These studies confounded number of relearning trials with ISI; that is, there was more relearning at longer ISIs. Some studies administered recognition tests without feedback during learning sessions (in some cases combined with recall tests; Burtt & Dobell, 1925; Spitzer, 1939; Welborn, 1933). Because these studies did not provide feedback, it is likely that no relearning occurred on the second and subsequent sessions for any item that elicited an error (see Pashler, Cepeda, Wixted, & Rohrer, 2005). Some studies (Simon, 1979; E. C. Strong, 1973; E. K. Strong, 1916) provided unlimited restudy time that did not include testing with feedback. For these studies, it is unclear how much information was acquired during relearning sessions, because testing was not performed, and it is possible that the amount of relearning and ISI were confounded. Some studies were conducted outside a laboratory setting. For example, the studies by Simon (1979) and E. C. Strong (1973) relied on participants reading unsolicited direct mail advertising. Regular adherence to the paradigm was unlikely, as the authors of these studies acknowledged.

In contrast to these confounded studies, other studies appear free of major confounds. Several experiments provided either learning to perfect performance on the first session or a fixed number of first-session learning trials, followed by a small, fixed number of study trials (with feedback) during the second session (Cepeda et al., 2005). These experiments equated, across conditions, the degree of initial learning (learning during the first session) and avoided any confound between subsequent learning (learning during the second session) and ISI. A number of studies had fixed (Ausubel, 1966; Childers & Tomasello, 2002; Edwards, 1917; Glenberg & Lehmann, 1980) restudy time, without feedback. Even though the amount of relearning that took place during the second session was not assessed, relearning was not confounded in these studies.

To provide some indication of the importance of these methodological issues, we examined the effect of ISI at similar retention intervals, comparing the studies we judged to be confounded with those we judged to be nonconfounded. There are seven experiments in five articles that used nonconfounded designs with ISIs and retention intervals of 1 day or more (Ausubel, 1966; Cepeda et al. 2005; Childers & Tomasello, 2002; Edwards, 1917; Glenberg & Lehmann, 1980). The Bahrick studies (Bahrick, 1979; Bahrick et al., 1993; Bahrick & Phelps, 1987), which confounded amount of relearning and ISI, showed similar patterns to Cepeda et al. (2005), Experiments 2a and 2b, which are unconfounded. The ideal ISI indicated in all these studies is 1 month or more, at retention intervals of 6 months or more. The Bahrick studies used far longer retention intervals than the Cepeda et al. study, making this comparison less than perfect. Burtt and Dobell (1925) and Spitzer (1939), who failed to provide relearning during relearning sessions for items that elicited errors, found that an ISI of 7–10 days was usually preferable to an ISI of 1–3 days, at retention intervals from 10–17 days. This contrasts with the unconfounded studies by Ausubel (1966); Cepeda et al., Experiment 1; and Glenberg and Lehmann (1980), who used similar retention intervals of 6–10

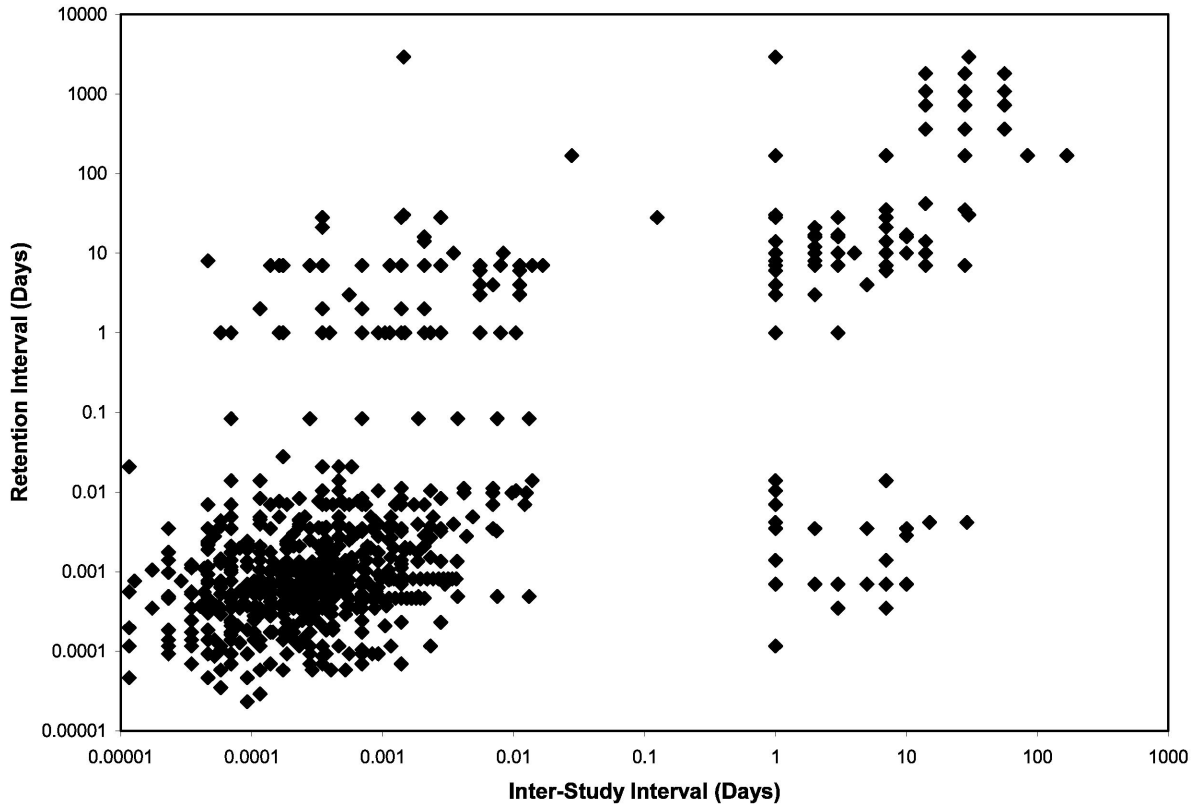

Figure 5. Scatter plot of interstudy interval by retention interval, for all studies in the absolute lag analyses.

days and who found that the ideal ISI was closer to 1–3 days than 7–10 days. Welborn (1933), who failed to provide relearning during relearning sessions for items that elicited errors, found effects similar to Cepeda et al.: In both studies, retention decreased as ISI increased beyond 1 day. However, Welborn used a retention interval of 28 days, whereas Cepeda et al. used a retention interval of 10 days. Two studies that used unlimited restudy time (Simon, 1979; E. C. Strong, 1973) are in line with similar unconfounded studies (i.e., Ausubel, 1966; Cepeda et al., 2005, Experiment 1; Glenberg & Lehmann, 1980), but one study that used unlimited restudy time (E. K. Strong, 1916) is not. Even with some inconsistencies between confounded and unconfounded experimental designs, we believe that our analyses of ISI and retention interval joint effects are not undermined by experimental design problems plaguing some of the experiments included in our analyses. Indeed, regardless of whether the confounded studies are excluded, the same basic conclusion would be drawn: Optimal ISI increases as retention interval increases.

Expanding Versus Fixed ISIs

It often has been suggested that when items are to be relearned on two or more occasions, memory can be maximized by relearning information at increasingly spaced (expanding) ISIs, as opposed to relearning at a fixed ISI (Bahrick & Phelps, 1987; Hollingworth, 1913; Kitson, 1921; Landauer & Bjork, 1978; Modigliani, 1967; Pyle, 1913). One intuitive version of this formulation says memory is best promoted when a learner undergoes tests that are as difficult as possible, while maintaining errorless performance. Only a few studies have empirically examined this issue, however, resulting in 22 comparisons of retention accuracy and 8 effect size comparisons. Independent samples *t* tests were used for analyses, as a conservative measure, as some studies were between subjects ($n = 7$) and others were within subject ($n = 11$).

Overall, expanding ISIs led to better performance than fixed intervals (see Table 8). Fifteen out of 18 studies used a paired associate learning task, and we did not detect any systematic differences related to type of task. Unfortunately, large standard errors, indicative of large between-study variability, make conclusions drawn from expanding versus fixed interval data necessarily tentative. Large between-study differences can be seen more dramatically by examining the empirical data from three different researchers, shown in Table 9. All three researchers used ISIs and retention intervals of at least 1 day. One researcher (Tsai, 1927) found better performance with expanding study intervals, one (Cull, 2000) found better performance with fixed study intervals, and one (Clark, 1928) found no difference between fixed and expanding intervals. In all three sets of studies, the average between-presentation ISI was the same for expanding and fixed ISIs, and retention intervals overlapped across studies; use of different ISIs and retention intervals does not explain differences between each set of studies. Any number of differences may explain these conflicting findings. One variable that might explain between-study differences is the presence of feedback. Expanding intervals might benefit performance when feedback is withheld, because expanding intervals minimize the chance of forgetting an item. (In the absence of feedback, forgetting an item usually causes

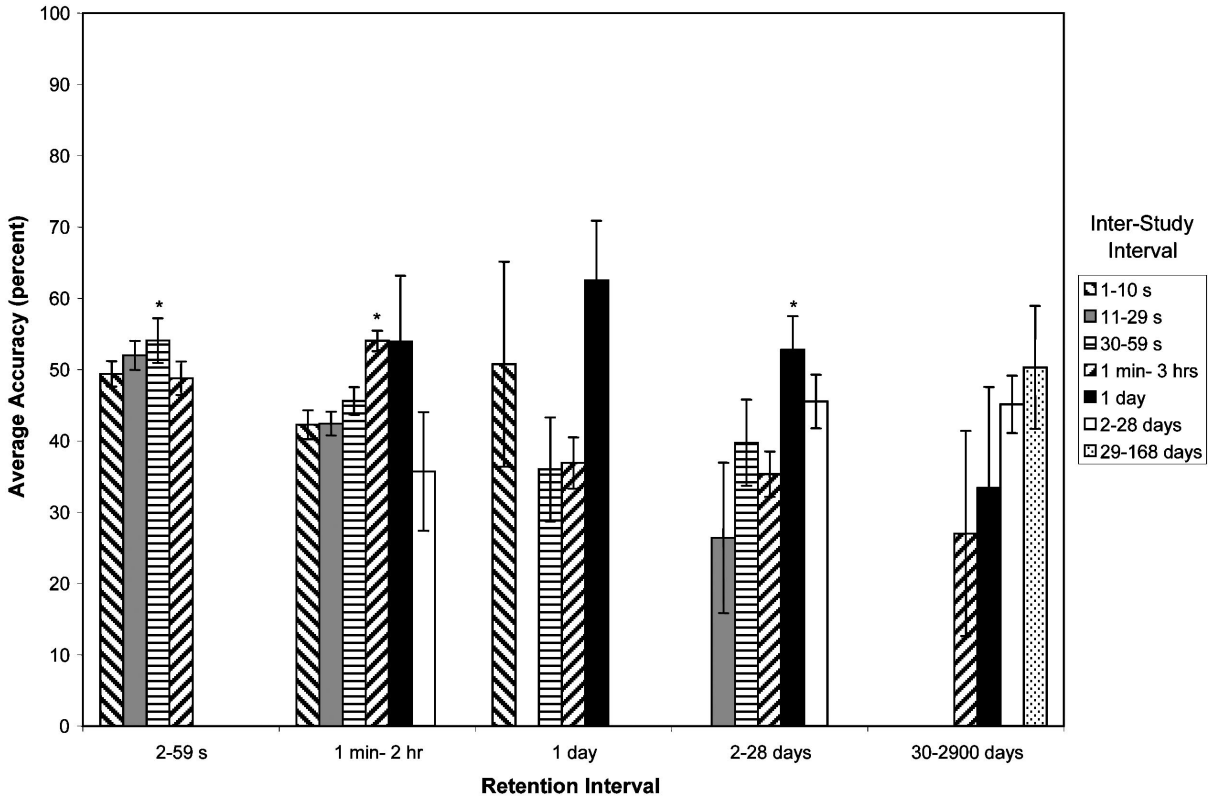

Figure 6. For all studies in the absolute lag analyses, accuracy, binned by interstudy interval (ISI) and retention interval and averaged across studies. When surrounded by ISI bins with lower accuracy values, the ISI bin showing the highest accuracy value at each retention interval bin is indicated with an asterisk. Error bars represent one standard error of the mean.

the item to be unrecoverable; see Pashler et al., 2005) This feedback hypothesis is supported by a single study (Cull, Shaughnessy, & Zechmeister, 1996). Unfortunately, the feedback hypothesis cannot be tested adequately with current data, because all three of the studies using ISIs and retention intervals longer than 1 day either provided testing with feedback (Cull, 2000) or provided a fixed amount of item restudy time (Clark, 1928; Cull, 2000; Tsai, 1927), which was functionally equivalent to providing feedback (because the entire to-be-learned item was present). With the exception of Cull et al. (1996) and Landauer and Bjork (1978), expanding interval studies that used retention intervals of less than 1 day (Cull, 1995; Foos & Smith, 1974; Hser & Wickens, 1989; Siegel & Misselt, 1984) all provided either a fixed amount of restudy time for each entire item or testing with feedback. We are left with inadequate evidence to support or refute the feedback hypothesis.

General Discussion

Although the distributed practice effect has spawned a large literature, prior meta-analyses (Donovan & Radosevich, 1999; Janiszewski et al., 2003; T. D. Lee & Genovese, 1988) failed to distinguish spacing effects (a single presentation, or a lag less than 1 s, vs. multiple presentations, or a lag of 1 s or more, of a given item; equal total study time for that item, whether in the spaced or massed condition) from lag effects (less vs. more time between study opportunities for a given item, when study opportunities for both the shorter and longer lag conditions are separated by 1 s or more). In the present review, this spacing versus lag distinction proved helpful in quantifying the relationship between level of retention, ISI, and retention interval. When participants learned individual items at two different points in time (spaced; lag of 1 s or more), equating total study time for each item, they recalled a greater percentage of items than when the same study time was nearly uninterrupted (massed; lag of less than 1 s). This improvement occurred regardless of whether the retention interval was less than 1 min or more than 1 month. In short, for the spacing effect proper, we failed to find any evidence that the effect is modulated by retention interval. At first blush, this conclusion might seem to suggest that students are wrong to believe that cramming immediately before an exam is an effective strategy to enhance performance on the exam. However, a few hours of cramming would typically involve repeated noncontiguous study of individual bits of information, rather than literal massing as examined in the studies noted. Furthermore, most advocates of cramming probably have in mind the comparison between studying immediately prior to the exam and studying days or weeks prior to the exam.

A different pattern of results was observed for increases in ISI beyond the massed condition (i.e., from a nonzero value to an even larger nonzero value). When ISI was increased, participants retained more information. However, for long ISIs, in proportion to

Table 5
Shorter and Longer Interstudy Interval (ISI) Range, Retention Interval Range, Percentage Correct at the Shorter and Longer ISI Range, and Satistical Analyses, for Absolute Lag Analyses

Shorter ISI range	Longer ISI range	Retention interval range	% correct at ISI range			Statistical analysis
			Shorter	Longer	*SEM*	
1–10 s	30–59 s	2–59 s	49.4	54.1	2.5	$t(162) = 1.4$, $p = .167$
30–59 s	1 min–3 hr	2–59 s	54.1	48.8	2.7	$t(90) = 1.3$, $p = .198$
1–10 s	1 min–3 hr	1 min–2 hr	42.3	54.0	1.7	$t(248) = 4.8$, $p < .001$
1 min–3 hr	2–28 days	1 min–2 hr	54.0	35.7	4.9	$t(161) = 3.4$, $p < .005$
30–59 s	1 day	1 day	36.0	62.5	7.8	$t(16) = 2.2$, $p < .05$
11–29 s	1 day	2–28 days	26.4	52.8	7.6	$t(21) = 2.5$, $p < .05$
1 day	2–28 days	2–28 days	52.8	45.5	4.2	$t(58) = 1.1$, $p = .270$
1 min–3 hr	29–168 days	30–2,900 days	27.0	50.3	11.5	$t(12) = 1.4$, $p = .180$

retention interval, further increases in ISI reduced accuracy. Thus, for a given retention interval, there was a nonzero value of ISI that optimized accuracy. (This is known as a nonmonotonic lag effect.) Moreover, the optimal ISI increased as retention interval increased. For instance, at retention intervals of less than 1 min, ISIs of less than 1 min maximized retention; at retention intervals of 6 months or more, ISIs of at least 1 month maximized retention. These results clearly show that a single ISI does not produce optimal retention across a wide range of retention intervals. The nonmonotonic effect of ISI upon retention and the dependency of optimal ISI upon retention interval both appear to characterize the literature as a whole, as well as a few well-known specific studies (e.g., Glenberg & Lehmann, 1980).

Some researchers have suggested, with little apparent empirical backing, that expanding ISIs improve long-term learning (Hollingworth, 1913; Kitson, 1921; Landauer & Bjork, 1978; Pyle, 1913); in contrast, some empirical studies (Cull, 1995, 2000; Foos & Smith, 1974) have found that expanding intervals are less effective than fixed spacing intervals. Our review of the evidence suggests that, in general, expanding intervals either benefit learning or produce effects similar to studying with fixed spacing. The literature offers examples of impaired performance with expanding intervals (Cull, 2000; Foos & Smith, 1974) and examples of expanding interval benefits (Cull et al., 1996; Hser & Wickens, 1989; Landauer & Bjork, 1978; Tsai, 1927). We found no obvious systematic differences between studies that do and do not show

Table 6
Number of Data Points and Studies, Unique Participants, and Data Points Using Paired Associate, List Recall, or Other Task Types, for Absolute Lag Analyses, by Retention Interval Range and Interstudy Interval (ISI) Range

Retention interval range	ISI range	No. of data points	No. of studies	No. of unique participants	No. of data points using		
					Paired associate tasks	List recall tasks	Other tasks
2–59 s	1–10 s	113	62	3,248	41	66	6
2–59 s	11–29 s	96	57	2,694	29	59	8
2–59 s	30–59 s	51	35	1,707	21	24	6
2–59 s	1 min–3 hr	41	20	1,152	14	27	0
1 min–2 hr	1–10 s	101	66	3,711	24	59	18
1 min–2 hr	11–29 s	84	76	4,773	25	53	6
1 min–2 hr	30–59 s	93	80	4,785	34	40	19
1 min–2 hr	1 min–3 hr	149	83	4,867	45	64	40
1 min–2 hr	1 day	11	8	222	2	5	4
1 min–2 hr	2–28 days	14	9	390	3	0	11
1 day	1–10 s	4	4	60	1	3	0
1 day	30–59 s	12	11	552	3	6	3
1 day	1 min–3 hr	30	19	1,100	12	16	2
1 day	1 day	6	6	83	0	4	2
2–28 days	11–29 s	5	5	190	4	0	1
2–28 days	30–59 s	8	5	267	0	4	4
2–28 days	1 min–3 hr	35	20	1,215	12	15	8
2–28 days	1 day	18	15	892	4	5	9
2–28 days	2–28 days	42	24	3,344	14	11	17
30–2,900 days	1 min–3 hr	4	3	53	4	0	0
30–2,900 days	1 day	5	3	54	4	0	1
30–2,900 days	2–28 days	15	4	175	13	0	2
30–2,900 days	29–168 days	10	3	84	10	0	0

Table 7
Final Test Performance for Long Interstudy Interval (ISI), Long Retention Interval Studies

Study	ISI (days)	Retention interval (days)	Final test performance (% correct)
Ausubel (1966)	1	6	43
	7	6	40
Bahrick (1979), Experiment 2	1	30	86
	30	30	95
Bahrick, Bahrick, Bahrick, & Bahrick (1993)	14	360	62
	28	360	67
	56	360	76
	14	720	55
	28	720	61
	56	720	67
	14	1,080	45
	28	1,080	62
	56	1,080	66
	14	1,800	36
	28	1,800	46
	56	1,800	60
Bahrick & Phelps (1987)	1	2,900	8
	30	2,900	15
Burtt & Dobell (1925), Experiment 2	3	10	22
	10	10	48
	3	16	16
	10	16	15
Burtt & Dobell (1925), Experiment 3	3	10	30
	10	10	55
	3	17	21
	10	17	25
Cepeda et al. (2005), Experiment 1	1	10	74
	2	10	69
	4	10	68
	7	10	69
	14	10	65
Cepeda et al. (2005), Experiment 2a	1	168	33
	7	168	47
	28	168	56
	84	168	43
	168	168	45
Cepeda et al. (2005), Experiment 2b	1	168	9
	7	168	14
	28	168	26
	84	168	19
	168	168	17
Childers & Tomasello (2002), Experiment 1	1	1	58
	3	1	58
	1	7	53
	3	7	53
Edwards (1917)	1	3	38
	2	3	19
	1	4	37
	5	4	32
Glenberg & Lehmann (1980), Experiment 2	1	7	32
	7	7	25
Simon (1979)	7	7	62
	28	7	43
	7	35	30
	28	35	31
Spitzer (1939)	1	14	36
	7	14	39
	14	14	39
E. C. Strong (1973)	7	7	16
	14	7	11
	28	7	11
E. K. Strong (1916)	1	28	13
	7	28	16
Welborn (1933)	1	28	72
	3	28	63

Table 8
Percentage Correct on the Final Recall Test for Expanding and Fixed Conditions, Number of Performance Differences and Studies, Total Number of Participants Summing Across All Study and Condition Combinations, and Statistical Analyses, for Expanding Versus Fixed Study Intervals

Retention interval	% Correct			No. of performance differences	No. of studies	No. of participants	Statistical analysis
	Expanding conditions	Fixed conditions	*SEM*				
1–59 s	91.0	91.0		1	1	24	
1 min–less than 10 min	49.8	48.9	5.6	10	8	580	$t(18) = 0.1$, $p = .91$
10 min–less than 1 day	77.8	70.0	11.5	4	3	614	$t(6) = 0.5$, $p = .65$
1 day							
2–7 days	66.3	59.5	10.9	4	3	185	$t(6) = 0.4$, $p = .68$
8–30 days	66.3	64.0	11.2	3	3	115	$t(4) = 0.1$, $p = .89$
31 days or more							
All retention intervals	62.0	58.6	4.6	22	18	1518	$t(42) = 0.5$, $p = .61$

expanding interval benefits, although one difference that might account for interstudy variability is the presence or absence of feedback. Given the practical import of multisession study (almost all learning takes place on more than two occasions), this topic clearly deserves further research.

Implications for Theories of Distributed Practice

Many theories purport to account for distributed practice effects, and little consensus has been achieved about the validity of these accounts. Although a thorough theoretical analysis of the distributed practice task is well beyond the scope of the present, relatively focused, review (for reviews of distributed practice, see Glenberg, 1979; Hintzman, 1974), it is of interest to examine how some of the principle conclusions reached in the present review might affect the credibility of some frequently discussed theories. We focus on four theories in detail, without in any way implying that other theories lack merit.

To date, theorists have failed to distinguish between spacing and lag effects. This makes it difficult to know how broadly theorists intended their theories to be applied. Theories often predict that spaced and massed items will be processed differently—for example, the inattention theory predicts that spaced items will receive greater attentional focus; the encoding variability theory predicts that spaced items will contain more interitem associations. (Massed items have associations only to the two immediately adjacent items, whereas spaced items have associations to at least three and usually four adjacent items. Spaced items have more associations because each spaced item is sandwiched between two items in the first session and sandwiched between two different items in the second session.) Because these and other theories are able to make differential predictions for spaced versus massed presentations, as well as for changes in lag, our theoretical discussion applies to both spacing and lag effects. In other words, our theoretical discussion applies to distributed practice effects, where distributed practice includes both spacing and lag effects.

Table 9
Percentage Correct on Final Test, for Fixed and Expanding Study Intervals, for Studies with a Retention Interval of at Least 1 Day

Study	ISI (days)	Retention interval (days)	Fixed study intervals (% correct)	Expanding study intervals (% correct)
Clark (1928)	2	21	63	63
Cull (2000), Exp. 3	2	3	98	84
Cull (2000), Exp. 4	2	8	89	82
Tsai (1927), Exp. 2	2	3	48	61
Tsai (1927), Exp. 2	2	7	36	46
Tsai (1927), Exp. 3	2	3	56	74
Tsai (1927), Exp. 3	2	17	40	54

Note. ISI = interstudy interval; Exp. = Experiment.

The first class of theoretical accounts that we discuss is deficient processing theory. Deficient processing theory is based on mechanisms that alter the amount of focus received by particular items. An example of deficient processing theory is the inattention theory (Hintzman, 1974). Inattention theory suggests that when the ISI is short, processing of the second presentation is reduced in quality and/or quantity: The learner pays less attention to something that is, by virtue of the short ISI, relatively more familiar. Deficient processing theory has struck many writers as offering an intuitively reasonable account of why massed presentations would produce inferior memory. The fact that massed presentations are normally inferior even when retention interval is very short, as noted above, certainly seems consistent with this account. This account also enjoys support from a study that suggests it is the trace of the second presentation, rather than the first, that is reduced when ISI is shorter than optimal (Hintzman, Block, & Summers, 1973).

Can deficient processing theory handle one of our meta-analysis's primary findings, the joint effects of ISI and retention interval? Suppose Study 1 yields a single memory trace, which is then further strengthened as a consequence of Study 2, and further suppose this trace is characterized by two parameters: the strength of the trace and its rate of decay. These two parameters are found in a number of functions used to describe forgetting, including the commonly preferred power law function described by Wixted and Ebbesen (1997). If Study 2 strengthens the trace without affecting its decay parameter, then even if the degree of strengthening is assumed to vary in some arbitrary fashion with ISI, there will have to be a single value of ISI that yields the strongest trace. This ISI would produce optimal later recall, regardless of how long the final test is delayed. Thus, this version of the deficient processing theory

is inconsistent with the effect of retention interval on optimal ISI, as seen in the present integrative review.

One could, of course, hypothesize that it is not just strength, but also decay rate, that is modified by Study 2 (making the account closer to suggestions by Pavlik & Anderson, 2003; Reed, 1977; and Wickelgren, 1972, discussed below), but this assumption is at odds with classic findings in the forgetting literature. That is, variations in the degree of attention paid to a study item appear to affect either the quantity or the quality of processing, but not both. Direct manipulations of the quantity of processing are known to have a large effect on the degree of learning (a proxy for strength) while having little or no effect on the rate of forgetting (Anderson, 2000; Underwood & Keppel, 1963; Wixted, 2004). Similarly, manipulating the quality of processing at encoding by manipulating depth of processing has a large effect on the degree of learning but a negligible effect on the rate of forgetting (McBride & Dosher, 1997). ISI, in contrast, has a large effect on the rate of forgetting. Specifically, as ISI increases, the rate of decay decreases, which is to say that longer ISIs produce more gradual forgetting curves. Nevertheless, it is conceivable that variations in attention affect the quality of processing in some other, as yet unspecified, way. If so, then the deficient processing theory may yet be able to accommodate our findings. In light of the available evidence, however, the effect of ISI on the rate of forgetting seems not to be an indirect result of the effect of that manipulation on attention.

Things become more complicated if one assumes that Study 1 and Study 2 produce two independent traces. One could, for example, suppose that the stronger is the trace resulting from Study 1 (call this Trace 1) at the time of Study 2, and the weaker is the trace formed from Study 2 (Trace 2). Once again, however, if it is assumed that Trace 1 strength affects the strength but not the decay rate of Trace 2, this independent-trace account also fails to explain the dependence of optimal ISI upon retention interval.

In summary, deficient processing theory appears to be threatened by complex joint effects of ISI and retention interval that were revealed in the literature, as documented in the present review. Although it would obviously be premature to say that all versions of the deficient processing account are falsified, the challenges appear substantial. (The deficient processing account confronts a separate difficulty in the finding that providing rewards for remembering does not reduce distributed practice effects; Hintzman, Summers, Eki, & Moore, 1975.)

A second widely discussed class of models is usually termed encoding variability theory (Glenberg, 1979; Melton, 1970). In the simplest versions of this account, traces stored when an item is studied represent the context in which the item is stored, as well as the item itself. Over time, the prevailing context is assumed to undergo random drift. As a result, the average distance between any prior context and the current context will increase with the passing of time. The account assumes that the shorter the distance between the context existing at retrieval and the context that existed at study, the greater the likelihood of retrieval success. Thus, as the ISI between Study 1 and Study 2 increases, the probability of later recall might grow, simply because it becomes more likely that the retrieval context will be similar to at least one of the study contexts. This can predict that the probability of later recall will grow as ISI increases, because it becomes more likely that the retrieval context will be similar to at least one of the study contexts.

Recent simulations (see Cepeda et al., 2005) demonstrate that a simple contextual drift mechanism—in conjunction with certain reasonable assumptions about the function relating similarity to retrieval probability—can readily produce distributed practice effects. Briefly, we created a simple model of encoding variability, based solely on contextual drift over time. Both context and time vary on a single dimension. Over time, location in one-dimensional contextual space changes and this change is either toward or away from the context at time x. Encouragingly, our simulations reveal that this simple version of encoding variability theory predicts both nonmonotonic effects of ISI and that the optimal ISI increases in a predictable fashion as retention interval increases (with the optimal ratio of ISI to retention interval decreasing as retention interval itself grows).

Encoding variability theory appears to encounter substantial problems when accounting for certain other findings (e.g., Bellezza, Winkler, & Andrasik, 1975; Dempster, 1987b). One potential problem for encoding variability theory comes from Ross and Landauer (1978), who showed that greater spacing between two instances of two different words presented at various list positions did not enhance the probability that the subject would later recollect either the first- or the second-presented item. In most versions of the encoding variability theory, one would expect such an enhancement for precisely the redundancy-related reasons noted above (see Raaijmakers, 2003, for a model of encoding variability that, according to its author, can be reconciled with Ross and Landauer's results). A second potential problem with encoding variability theory is when participants are deliberately induced to encode items in a more variable fashion, this often fails to produce a later recall benefit or fails to modulate the distributed practice effect (Dempster, 1987a; Hintzman & Stern, 1977; Maki & Hasher, 1975; Maskarinec & Thompson, 1976; McDaniel & Pressley, 1984; Postman & Knecht, 1983).

A third explanation for the distributed practice effect is termed consolidation theory (Wickelgren, 1972). Upon the second presentation of a repeated item, consolidation theory proposes that a new (second) trace is formed that inherits the state of consolidation of the first occurrence of that item. If the ISI is 1 week, more consolidation into long-term memory will have occurred than if the ISI is 1 day, and the second trace will inherit this higher state of consolidation. If the delay is too long, say 1 year, there will be no initial memory trace whose consolidation state can be inherited, and thus retention of that item will be lowered. This theory, as well as related accounts proposed by Pavlik and Anderson (2003) and Reed (1977), quite directly predicts that, for a given retention interval, ISI varies nonmonotonically; it may or may not also predict that optimal ISI increases monotonically with retention interval.

One experimental result that appears to undercut consolidation theory is the finding of Hintzman et al. (1973), which suggests that learning produced by Study 2, rather than learning produced by Study 1, is decreased when the Study 2 presentation follows closely after the Study 1 presentation (see Murray, 1983, for arguments that this finding may not be definitive). If Study 1 processing were interrupted, as purported in consolidation theory, then Study 1 and not Study 2 learning should be decreased.

Study-phase retrieval theory (Braun & Rubin, 1998; Murray, 1983; Thios & D'Agostino, 1976) provides a fourth explanation of the distributed practice effect. In this theory, the second (restudy) presentation serves as a cue to recall the memory trace of the first

presentation. This is similar to consolidation theory, but unlike in consolidation theory, consolidation of the first-presentation memory trace is not interrupted. Study-phase retrieval is supported by empirical evidence: A lag effect is found when retrieval of the first presentation is required (Thios & D'Agostino, 1976); in contrast, no lag effect is found when retrieval is not required. Notably, interrupting or otherwise diminishing study-phase retrieval can eliminate the distributed practice effect (Thios & D'Agostino, 1976). The mechanism(s) by which retrieval of the first-presentation trace helps later retrieval has been left open to interpretation: Sources of benefit may include increased contextual associations or strengthened first-presentation traces. As in consolidation theory, if the first-presentation memory trace cannot be retrieved, then later retrieval will be less likely; thus, study-phase retrieval theory predicts nonmonotonic lag effects. It is unclear whether study-phase retrieval theory predicts that optimal ISI increases monotonically with retention interval.

In summary, the findings gleaned in the present quantitative synthesis appear to have a significant bearing on the four potential theories of the distributed practice effect discussed here. At least on the basis of our preliminary analysis, study-phase retrieval, consolidation, and encoding variability theories survive as candidate distributed practice theories, whereas deficient processing theory does not readily survive. Notably, only encoding variability theory has been shown, through mathematical modeling, to produce increases in optimal ISI as retention interval increases. It remains unclear whether consolidation and/or study-phase retrieval theory can produce this effect and whether these results can be reconciled with the empirical challenges that have been arrayed against them, as noted above. Further analytic work is needed to explore in more detail the relationship between potential theories of distributed practice and the finding that optimal ISI increases as retention interval increases.

Educational Implications of Findings

A primary goal of almost all education is to teach material so that it will be remembered for an extended period of time, on the order of at least months and, more often, years. The data described here reaffirm the view (expressed most forcefully by Bahrick, 2005, and Dempster, 1988) that separating learning episodes by a period of at least 1 day, rather than concentrating all learning into one session, is extremely useful for maximizing long-term retention. Every study examined here with a retention interval longer than 1 month (Bahrick, 1979; Bahrick, et al., 1993; Bahrick & Phelps, 1987; Cepeda et al., 2005) demonstrated a benefit from distribution of learning across weeks or months, as opposed to learning across a 1-day interval; learning within a single day impaired learning, compared with a 1-day interval between study episodes; learning at one single point in time impaired learning, compared with a several-minute interval between study episodes. The average observed benefit from distributed practice (over massed practice) in these studies was 15%, and it appeared to hold for children (Bloom & Shuell, 1981; Childers & Tomasello, 2002; Edwards, 1917; Fishman, Keller, & Atkinson, 1968; Harzem, Lee, & Miles, 1976) as well as adults. After more than a century of research on spacing, much of it motivated by the obvious practical implications of the phenomenon, it is unfortunate that we cannot say with certainty how long the ISI should be to optimize long-term retention. The present results suggest that the optimal ISI increases as the duration over which information needs to be retained increases. For most practical purposes, this retention interval will be months or years, so the optimal ISI will likely be well in excess of 1 day. Obviously, there is a need for much more detailed study on this point, despite the time-consuming nature of such studies. One question of particular practical interest is whether ISIs that are longer than the optimal ISI produce large decrements in retention or only minor ones. If they produce only minor decrements in retention, then a simple principle "seek to maximize lag wherever possible" may be workable. On the other hand, if these decrements are substantial, then a serious consideration of the expected duration over which memory access will be needed may often be needed if one is to maximize the efficiency of learning.

Analysis Limitations

The present analysis is subject to many of the same limitations present in all meta-analyses (for discussion, see Hedges & Olkin, 1985; Hunter & Schmidt, 1990). For example, there is no way to accurately calculate the number of studies with null findings (i.e., a lack of distributed practice effect), because many studies never reach publication. This "file drawer problem" (Rosenthal, 1979) reflects the reluctance of journals to publish null findings. Hunter and Schmidt (1990) point out that the file drawer problem tends to be a nonissue when large effect sizes are identified, as in the present analysis, because of the enormous ratio of unpublished to published data that would be needed to invalidate a large effect size.

Limitations of Currently Available Data

As noted above, new studies are sorely needed to clarify the effects of interstudy and retention intervals that are educationally relevant, that is, on the order of weeks, months, or years. It is clear from existing studies that the distribution of a given amount of study time over multiday periods produces better long-term retention than study over a few-minute period, but it is unclear how quickly retention drops off when intervals exceed the optimal ISI. If the field of learning and memory is to inform educational practice, what is evidently needed is much less emphasis on convenient single-session studies and much more research with meaningful retention intervals (see Bahrick, 2005, for similar comments).

The effects of nonconstant (i.e., expanding or contracting) learning schedules on retention are still poorly understood. Expanding study intervals rarely seem to produce much harm for recall after long delays, but there is insufficient data to say whether they help. This has not stopped some software developers from assuming that expanding study intervals work better than fixed intervals. For example, Woźniak and Gorzelańczyk (1994; see also SuperMemo World, n.d.) offered a "universal formula" designed to space repetitions at an interval that will produce 95% retention, based on Bahrick and Phelps's (1987) proposal that the ideal spacing interval is the longest ISI before items are forgotten.

We sometimes found it necessary to focus on change in accuracy as a measure, instead of the more traditional effect size measure, because the variance data necessary to compute effect size were lacking in most published results in this area. It was very encouraging to observe that results differed little depending upon

whether accuracy difference or effect size was examined. Future research in the area of distributed practice should report the sample size, means, and standard deviations for each ISI data point, even in cases of no significant difference, so that effect size can be calculated in future meta-analyses (American Psychological Association, 2001). As well, it would be useful if researchers reported pairwise correlations between ISIs, so that dependence between responses can be corrected, whenever the design is within subjects.

Almost all distributed practice data in our analysis (85%) are based on performance of young adults (see Table 10). Although most studies using children show a distributed practice effect, there simply is insufficient data to make strong claims about the similarity between children's and adults' responses to distributed practice, when retention interval is 1 day or longer. Until empirical data examining the distributed practice effect in children are collected, using retention intervals of months or years and ISIs of days or months (no usable data meeting these criteria currently exist, to our knowledge), we cannot say for certain that children's long-term memory will benefit from distributed practice.

Summary

More than 100 years of distributed practice research have demonstrated that learning is powerfully affected by the temporal distribution of study time. More specifically, spaced (vs. massed) learning of items consistently shows benefits, regardless of retention interval, and learning benefits increase with increased time lags between learning presentations. On the other hand, it seems clear that once the interval between learning sessions reaches some relatively long amount of time, further increases either have no effect upon or decrease memory as measured in a later test. The magnitude of the observed distributed practice benefit depends on the joint effects of ISI and retention interval; retention interval influences the peak of this function. Distributing learning across different days (instead of grouping learning episodes within a single day) greatly improves the amount of material retained for sizable periods of time; the literature clearly suggests that distributing practice in this way is likely to markedly improve students' retention of course material. Results also show that despite the sheer volume of the distributed practice literature, some of the most practically important questions remain open, including magnitude of the drop-off produced by use of a supraoptimal ISI, the relative merits of expanding (as compared with uniformly spacing) learning sessions, and the range of ISI values needed to promote memory durability over the range of time to which educators typically aspire. We have little doubt that relatively expensive and time-consuming studies involving substantial retention intervals will need to be carried out if practical benefits are to be wrung from distributed practice research; it is hoped that the present review will help researchers to pinpoint where that effort might be the most useful and illuminating.

Table 10
Number of Performance Differences and Effect Sizes, by Age Group

Age group	No. of performance differences	No. of effect sizes
Preschool	29	13
Elementary school	26	3
Junior high school	29	3
High school	14	6
Young adult (18–35)	714	97
Middle-aged adult (36–60)	2	0
Older adult (61+)	11	2
Mixed adult (18+)	14	9

Note. Ages are in years.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Actkinson, T. R. (1977). The effects of structural, presentation, and response variables on storage and recall of simple prose materials (Doctoral dissertation, Texas Christian University, 1976). *Dissertation Abstracts International, 37,* 5388.

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin, 101,* 41–74.

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Anderson, J. R. (2000). *Learning and memory: An integrated approach* (2nd ed.). New York: Wiley.

*Austin, S. D. M. (1921). A study of logical memory. *American Journal of Psychology, 32,* 370–403.

*Ausubel, D. P. (1966). Early versus delayed review in meaningful learning. *Psychology in the Schools, 3,* 195–198.

*Azerrad, J., & Kennedy, W. A. (1969). Short-term recall of CVC trigrams. *Psychological Reports, 24,* 698.

*Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General, 108,* 296–308.

Bahrick, H. P. (2005). The long-term neglect of long-term memory: Reasons and remedies. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications: Decade of behavior* (pp. 89–100). Washington, DC: American Psychological Association.

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

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

*Balota, D. A., Duchek, J. M., & Paullin, R. (1989). Age-related differences in the impact of spacing, lag, and retention interval. *Psychology and Aging, 4,* 3–9.

*Batchelder, W. H., & Riefer, D. M. (1980). Separation of storage and retrieval factors in free recall of clusterable pairs. *Psychological Review, 87,* 375–397.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41,* 257–278.

*Bellezza, F. S., Winkler, H. B., & Andrasik, F., Jr. (1975). Encoding processes and the spacing effect. *Memory & Cognition, 3,* 451–457.

*Bellezza, F. S., & Young, D. R. (1989). Chunking of repeated events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 990–997.

*Bird, C. P., Nicholson, A. J., & Ringer, S. (1978). Resistance of the spacing effect to variations in encoding. *American Journal of Psychology, 91,* 713–721.

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

*Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research, 74,* 245–248.

*Borkowski, J. G. (1967). Distributed practice in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 6,* 66–72.

*Braun, K. A. (1995). A retrieval model of the spacing effect (Doctoral dissertation, Duke University, 1994). *Dissertation Abstracts International, 55,* 3615.

*Braun, K., & Rubin, D. C. (1998). The spacing effect depends on an encoding deficit, retrieval, and time in working memory: Evidence from once-presented words. *Memory, 6,* 37–65.

*Bray, J. F., Robbins, D., & Witcher, W. B., Jr. (1976). Encoding variability theory and the spacing effect in associate learning. *Memory & Cognition, 4,* 548–552.

*Bregman, A. S. (1967). Distribution of practice and between-trials interference. *Canadian Journal of Psychology, 21,* 1–14.

Briggs, D. C. (2005). Meta-analysis: A case study. *Evaluation Review, 29,* 87–127.

*Brown, J., & Huda, M. (1961). Response latencies produced by massed and spaced learning of a paired-associates list. *Journal of Experimental Psychology, 61,* 360–364.

*Brown, W. (1924). Effects of interval on recall. *Journal of Experimental Psychology, 7,* 469–474.

*Burtt, H. E., & Dobell, E. M. (1925). The curve of forgetting for advertising material. *Journal of Applied Psychology, 9,* 5–21.

*Carter, J. F., & Carrier, C. (1976). Prose organization and recall. *Contemporary Educational Psychology, 1,* 329–345.

*Cepeda, N. J., Mozer, M. C., Coburn, N., Rohrer, D., Wixted, J. T., & Pashler, H. (2005). *Optimizing distributed practice: Theoretical analysis and practical implications.* Manuscript submitted for publication.

*Cermak, L. S., Verfaellie, M., Lanzoni, S., Mather, M., & Chase, K. A. (1996). Effect of spaced repetitions on amnesia patients' recall and recognition performance. *Neuropsychology, 10,* 219–227.

*Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 389–396.

*Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology, 38,* 967–978.

*Ciccone, D. S., & Brelsford, J. W. (1976). Spacing repetitions in paired-associate learning: Experimenter versus subject control. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 446–455.

*Clark, B. E. (1928). The effect upon retention of varying lengths of study periods and rest intervals in distributed learning time. *Journal of Educational Psychology, 19,* 552–559.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cortina, J. M., & Nouri, H. (2000). *Effect sizes for ANOVA designs.* Thousand Oaks, CA: Sage.

Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Erlbaum.

*Cuddy, L. J., & Jacoby, L. L. (1982). When forgetting helps memory: An analysis of repetition effects. *Journal of Verbal Learning and Verbal Behavior, 21,* 451–467.

*Cull, W. L. (1995). How and when should information be restudied? (Doctoral dissertation, Loyola University Chicago, 1995). *Dissertation Abstracts International, 56,* 549.

*Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14,* 215–235.

*Cull, W. L., Shaughnessy, J. J., & 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.

*D'Agostino, P. R. (1974). Repetition and recall in an incidental-memory paradigm. *Canadian Journal of Psychology, 28,* 468–473.

*D'Agostino, P. R., & DeRemer, P. (1972). Item repetition in free and cued recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 54–58.

*D'Agostino, P. R., & DeRemer, P. (1973). Repetition effects as a function of rehearsal and encoding variability. *Journal of Verbal Learning and Verbal Behavior, 12,* 108–113.

D'Amico, E. J., Neilands, T. B., & Zambarano, R. (2001). Power analysis for multivariate and repeated measures designs: A flexible approach using the SPSS MANOVA procedure. *Behavior Research Methods, Instruments, and Computers, 33,* 479–484.

*Dellarosa, D., & Bourne, L. E., Jr. (1985). Surface form and the spacing effect. *Memory & Cognition, 13,* 529–537.

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

Dempster, F. N. (1987b). Time and the production of classroom learning: Discerning implications from basic research. *Educational Psychologist, 22,* 1–21.

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

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review, 1,* 309–330.

*DeRemer, P., & D'Agostino, P. R. (1974). Locus of distributed lag effect in free recall. *Journal of Verbal Learning and Verbal Behavior, 13,* 167–171.

Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect. *Journal of Applied Psychology, 84,* 795–805.

Doré, L. R., & Hilgard, E. R. (1938). Spaced practice as a test of Snoddy's two processes in mental growth. *Journal of Experimental Psychology, 23,* 359–374.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170–177.

*Durgunoğlu, A. Y., Mir, M., & Ariño-Martí, S. (1993). Effects of repeated readings on bilingual and monolingual memory for text. *Contemporary Educational Psychology, 18,* 294–317.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger, C. E. Bussenius, & E. R. Hilgard, Trans.). New York: Dover Publications. (Original work published 1885)

*Edwards, A. S. (1917). The distribution of time in learning small amounts of material. In *Studies in Psychology Contributed by Colleagues and Former Students of Edward Bradford Titchener* (pp. 209–213). Worcester, MA: Louis N. Wilson.

*Elmes, D. G., Chapman, P. F., & Selig, C. W. (1984). Role of mood and connotation in the spacing effect. *Bulletin of the Psychonomic Society, 22,* 186–188.

*Elmes, D. G., Dye, C. J., & Herdelin, N. J. (1983). What is the role of affect in the spacing effect? *Memory & Cognition, 11,* 144–151.

*Elmes, D. G., Greener, W. I., & Wilkinson, W. C. (1972). Free recall of items presented after massed- and distributed-practice items. *American Journal of Psychology, 85,* 237–240.

*Elmes, D. G., Sanders, L. W., & Dovel, J. C. (1973). Isolation of massed- and distributed-practice items. *Memory & Cognition, 1,* 77–79.

*Evans, J. D. (1975). The lag effect in free recall: Differential encoding or differential rehearsal? (Doctoral dissertation, Iowa State University, 1974). *Dissertation Abstracts International, 35,* 4142.

*Fishman, E. J., Keller, L., & Atkinson, R. C. (1968). Massed versus distributed practice in computerized spelling drills. *Journal of Educational Psychology, 59,* 290–296.

*Folarin, B. A. (1983). The effect of spacing category members on children's memory. *Journal of Psychology, 114,* 167–177.

*Foos, P. W., & Smith, K. H. (1974). Effects of spacing and spacing patterns in free recall. *Journal of Experimental Psychology, 103,* 112–116.

*Gagné, R. M. (1969). Context, isolation, and interference effects on the retention of fact. *Journal of Educational Psychology, 60,* 408–414.

*Gartman, L. M., & Johnson, N. F. (1972). Massed versus distributed repetition of homographs: A test of the differential-encoding hypothesis. *Journal of Verbal Learning and Verbal Behavior, 11,* 801–808.

Gibbons, R. D., Hedeker, D. R., & Davis, J. M. (1993). Estimation of effect size from a series of experiments involving paired comparisons. *Journal of Educational Statistics, 18,* 271–279.

*Gilliland, T. R. (2000). Intervening tasks and the spacing effect: Theoretical and practical implications (Doctoral dissertation, Texas A&M University, 1999). *Dissertation Abstracts International, 60,* 4920.

*Glanzer, M. (1969). Distance between related words in free recall: Trace of the STS. *Journal of Verbal Learning and Verbal Behavior, 8,* 105–111.

*Glanzer, M., & Duarte, A. (1971). Repetition between and within languages in free recall. *Journal of Verbal Learning and Verbal Behavior, 10,* 625–630.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

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

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

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

*Glenberg, A. M., & Lehmann, T. S. (1980). Spacing repetitions over 1 week. *Memory & Cognition, 8,* 528–538.

*Glenberg, A. M., & Melton, A. W. (1978). *An analysis of the free recall spacing effect using the overt rehearsal technique.* Unpublished manuscript, University of Wisconsin—Madison and University of Michigan, Ann Arbor.

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

Gleser, L. J., Olkin, I. (1994). Stochastically dependent effect sizes. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261–281). New York: Russell Sage Foundation.

*Gordon, K. (1925). Class results with spaced and unspaced memorizing. *Journal of Experimental Psychology, 7,* 337–343.

*Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 371–377.

*Greene, R. L. (1990). Spacing effects on implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1004–1011.

Greene, R. L. (1992). *Human memory: Paradigms and paradoxes.* Hillsdale, NJ: Erlbaum.

*Hall, J. W. (1992). Unmixed effects of spacing on free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 608–614.

*Hall, J. W., Smith, T. A., Wegener, S. L., & Underwood, B. J. (1981). Rate and frequency as determinants of learning with complete and discrete list presentation. *Memory & Cognition, 9,* 360–367.

*Harzem, P., Lee, I., & Miles, T. R. (1976). The effects of pictures on learning to read. *British Journal of Educational Psychology, 46,* 318–322.

*Hayes-Roth, B. (1976). *Effects of repetitions and questions at varying lags during self-paced learning from text* (Rand Paper P-5925). Santa Monica, CA: Rand Corporation.

*Hecht, D. J. (1982). Another ten years of massed practice on distributed practice (Doctoral dissertation, University of Texas at Austin, 1982). *Dissertation Abstracts International, 43,* 899.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium* (pp. 77–97). Potomac, MD: Erlbaum.

Hintzman, D. L., Block, R. A., & Summers, J. J. (1973). Modality tags and memory for repetitions: Locus of the spacing effect. *Journal of Verbal Learning and Verbal Behavior, 12,* 229–238.

*Hintzman, D. L., & Stern, L. D. (1977). Failure to confirm Elmes, Greener, and Wilkinson's finding on the spacing effect. *American Journal of Psychology, 90,* 489–497.

Hintzman, D. L., Summers, J. J., Eki, N. T., & Moore, M. D. (1975). Voluntary attention and the spacing effect. *Memory & Cognition, 3,* 576–580.

*Hohn, R. L. (1972). *Effects of massed vs. distributed practice and word frequency on young children's free recall* (Rep. No. PS005604). Retrieved March 20, 2006, from ERIC database: http://eric.ed.gov/

Hollingworth, H. L. (1913). Advertising and selling: Principles of appeal and response. New York: D. Appleton.

*Houston, J. P. (1966). List differentiation and distributed practice. *Journal of Experimental Psychology, 72,* 477–478.

*Houston, J. P., & Reynolds, J. H. (1965). First-list retention as a function of list differentiation and second-list massed and distributed practice. *Journal of Experimental Psychology, 69,* 387–392.

*Hovland, C. I. (1940). Experimental studies in rote-learning theory: VI. Comparison of retention following learning to same criterion by massed and distributed practice. *Journal of Experimental Psychology, 26,* 568–587.

*Hser, Y.-I., & Wickens, T. D. (1989). The effects of the spacing of test trials and study trials in paired-association learning. *Educational Psychology, 9,* 99–120.

Hull, C. L. (1943). *Principles of behavior.* New York: Appleton-Century.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

Irion, A. L. (1966). A brief history of research on the acquisition of skill. In E. A. Bilodeau (Ed.), *Acquisition of skill* (pp. 1–46). New York: Academic Press.

*Izawa, C. (1970). List versus items in distributed practice in paired-associate learning. *Proceedings of the Annual Convention of the American Psychological Association, 5,* 87–88.

*Izawa, C. (1971). Massed and spaced practice in paired-associate leaning: List versus item distributions. *Journal of Experimental Psychology, 89,* 10–21.

*Izawa, C. (1979). Toward a comprehensive theory of variable performance differences and differential spaced practice effects between anticipation and study-test methods. *Journal of General Psychology, 100,* 63–83.

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

Janiszewski, C., Noel, H., & Sawyer, A. G. (2003). A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *Journal of Consumer Research, 30,* 138–149.

*Jensen, T. D., & Freund, J. S. (1981). Persistence of the spacing effect in incidental free recall: The effect of external list comparisons and intertask correlations. *Bulletin of the Psychonomic Society, 18,* 183–186.

Johnson, B. T., & Eagly, A. H. (2000). Quantitative synthesis of social psychological research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 496–528). New York: Cambridge University Press.

*Johnston, W. A., Coots, J. H., & Flickinger, R. G. (1972). Controlled semantic encoding and the effect of repetition lag on free recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 784–788.

*Johnston, W. A., & Uhl, C. N. (1976). The contributions of encoding effort and variability to the spacing effect on free recall. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 153–160.

Jost, A. (1897). Die Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen [The strength of associations in their dependence on the distribution of repetitions]. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane, 14,* 436–472.

*Kahana, M. J., & Greene, R. L. (1993). Effects of spacing on memory for homogenous lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 159–162.

*Kahana, M. J., Kamins, B. J., & Howard, M. W. (2003). *Spacing and lag effects are obtained in free recall of pure lists.* Unpublished manuscript.

*Keppel, G. (1964). Facilitation in short- and long-term retention of paired associates following distributed practice in learning. *Journal of Verbal Learning and Verbal Behavior, 3,* 91–111.

*Keppel, G. (1967). A reconsideration of the extinction-recovery theory. *Journal of Verbal Learning and Verbal Behavior, 6,* 476–487.

*Kitao, N., & Inoue, T. (1998). The effects of spaced repetition on explicit and implicit memory. *Psychologia, 41,* 114–119.

Kitson, H. D. (1921). *The mind of the buyer: A psychology of selling.* New York: MacMillan.

*Kraft, R. N., & Jenkins, J. J. (1981). The lag effect with aurally presented passages. *Bulletin of the Psychonomic Society, 17,* 132–134.

*Krug, D., Davis, B., & Glover, J. A. (1990). Massed versus distributed repeated reading: A case of forgetting helping recall? *Journal of Educational Psychology, 82,* 366–371.

*Landauer, T. K. (1967). Interval between item repetitions and free recall memory. *Psychonomic Science, 8,* 439–440.

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

*Landauer, T. K., & Eldridge, L. (1967). Effect of tests without feedback and presentation-test interval in paired-associate learning. *Journal of Experimental Psychology, 75,* 290–298.

*Landauer, T. L., & Rubin, N. (1966). Spacing of item repetitions in paired-associate learning. *Proceedings of the Annual Convention of the American Psychological Association,* 91–92.

*Lee, C. L. (1976). Repetition and acoustic contrast in short-term memory for letter sequences. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 695–704.

Lee, T. D., & Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport, 59,* 277–287.

*Leng, N. R. C., & Parkin, A. J. (1989). Aetiological variation in the amnesic syndrome: Comparisons using the Brown-Peterson task. *Cortex, 25,* 251–259.

*Madigan, S. A. (1970). Intraserial repetition and coding processes in free recall (Doctoral dissertation, University of Toronto, 1968). *Dissertation Abstracts International, 31,* 933–934.

*Madsen, M. C. (1963). Distribution of practice and level of intelligence. *Psychological Reports, 13,* 39–42.

*Madsen, M. C. (1966). Individual differences and temporal factors in memory consolidation. *American Journal of Mental Deficiency, 71,* 501–507.

Maki, R. H., & Hasher, L. (1975). Encoding variability: A role in immediate and long term memory? *American Journal of Psychology, 88,* 217–231.

*Marmurek, H. H. C., Holt, P. D., & Coe, K. (1978). Presentation mode and repetition effects in free recall. *American Journal of Psychology, 91,* 483–490.

*Maskarinec, A. S., & Thompson, C. P. (1976). The within-list distributed practice effect: Tests of the varied context and varied encoding hypotheses. *Memory & Cognition, 4,* 741–746.

*Mathews, R. C., & Tulving, E. (1973). Effects of three types of repetition on cued and noncued recall of words. *Journal of Verbal Learning and Verbal Behavior, 12,* 707–721.

*McAllister, E. W. C. (1973). Effects of type of recall test and spaced presentations of stimulus terms, response terms and stimulus-response terms on retention in paired-associate learning. *Psychological Reports, 33,* 307–311.

*McAllister, E. W. C. (1976). Effects of associative reaction time and spaced presentations of stimulus-test items, response-test items, and stimulus-response repetitions on retention in paired associate learning. *Bulletin of the Psychonomic Society, 7,* 205–207.

*McAllister, E. W. C., & Ley, R. (1972). Effects of spaced presentations of stimulus terms, response terms, and stimulus-response pairs on retention in paired-associate learning. *Journal of Experimental Psychology, 94,* 68–73.

McBride, D. M., & Dosher, B. A. (1997). A comparison of forgetting in an implicit and explicit memory task. *Journal of Experimental Psychology: General, 126,* 371–392.

*McCormack, P. D., & Carboni, N. L. (1973). Lag invariance with forced encodings in free recall. *Canadian Journal of Psychology, 27,* 144–151.

*McCullough, M. H. (1980). The role of repetition and context in the free recall of pictures and words. *Dissertation Abstracts International, 40* (7), 3452B. (UMI No. AAT 8000423)

McDaniel, M. A., & Pressley, M. (1984). Putting the keyword method in context. *Journal of Educational Psychology, 76,* 598–609.

*McFarland, C. E., Jr., Rhodes, D. D., & Frey, T. J. (1979). Semantic-feature variability and the spacing effect. *Journal of Verbal Learning and Verbal Behavior, 18,* 163–172.

McGeoch, J. A., & Irion, A. L. (1952). The distribution of practice and reminiscence. In *The psychology of human learning* (2nd ed., pp. 138–193). New York: David McKay.

*Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 596–606.

*Modigliani, V. (1967). Effects on a later recall by delaying initial recall. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 609–622.

Morris, S. B. (2000). Distribution of the standardized mean change effect size for meta-analysis on repeated measures. *British Journal of Mathematical and Statistical Psychology, 53,* 17–29.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7,* 105–125.

Moss, V. D. (1996). The efficacy of massed versus distributed practice as a function of desired learning outcomes and grade level of the student (Doctoral dissertation, Utah State University, 1995). *Dissertation Abstracts International, 56,* 5204.

*Murray, J. T. (1983). Spacing phenomena in human memory: A study-phase retrieval interpretation (Doctoral dissertation, University of California, Los Angeles, 1982). *Dissertation Abstracts International,* 43, 3058.

*Nelson, T. O. (1977). Repetition and depth of processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 151–171.

*Paivio, A. (1974). Spacing of repetitions in the incidental and intentional free recall of pictures and words. *Journal of Verbal Learning and Verbal Behavior, 13,* 497–511.

*Pashler, H., & Cepeda, N. J. (2003). [Within-subjects replication of Glenberg and Lehmann (1980), Experiment 2]. Unpublished raw data.

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 3–8.

*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.

*Patten, E. F. (1938). The influence of distribution of repetitions on certain rote learning phenomena. *Journal of Psychology, 5,* 359–374.

*Pavlik, P. I., & Anderson, J. R. (2003). An ACT-R model of the spacing effect. In F. Detje, D. Dorner, & H. Schaub (Eds.), *Proceedings of the Fifth International Conference of Cognitive Modeling* (pp. 177–182). Bamberg, Germany: Universitats-Verlag Bamberg.

*Perruchet, P. (1989). The effect of spaced practice on explicit and implicit memory. *British Journal of Psychology, 80,* 113–130.

*Peterson, L. R., & Gentile, A. (1965). Proactive interference as a function of time between tests. *Journal of Experimental Psychology, 70,* 473–478.

*Peterson, L. R., Hillner, K., & Saltzman, D. (1962). Supplementary report: Time between pairings and short-term retention. *Journal of Experimental Psychology, 64,* 550–551.

*Peterson, L. R., Saltzman, D., Hillner, K., & Land, V. (1962). Recency and frequency in paired-associate learning. *Journal of Experimental Psychology, 63,* 396–403.

*Peterson, L. R., Wampler, R., Kirkpatrick, M., & Saltzman, D. (1963). Effect of spacing presentations on retention of a paired associate over short intervals. *Journal of Experimental Psychology, 66,* 206–209.

Postman, L., & Knecht, K. (1983). Encoding variability and retention. *Journal of Verbal Learning and Verbal Behavior, 22,* 133–152.

*Potts, G. R., & Shiffrin, R. M. (1970). Repetitions, blank trials, and the vonRestorff effect in free recall memory. *Journal of Experimental Psychology, 86,* 128–130.

Pyle, W. H. (1913). Economical learning. *Journal of Educational Psychology, 4,* 148–158.

Raaijmakers, J. G. W. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science, 27,* 431–452.

*Rea, C. P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning, 4,* 11–18.

*Rea, C. P., & Modigliani, V. (1987). The spacing effect in 4- to 9-year-old children. *Memory & Cognition, 15,* 436–443.

Reed, A. V. (1977). Quantitative prediction of spacing effects in learning. *Journal of Verbal Learning and Verbal Behavior, 16,* 693–698.

*Robbins, D., & Bray, J. F. (1974). The spacing effect and the A-B, A-C paradigm: Evidence for retroactive facilitation. *Journal of Experimental Psychology, 103,* 420–425.

*Robinson, E. S. (1921). The relative efficiencies of distributed and concentrated study in memorizing. *Journal of Experimental Psychology, 4,* 327–343.

*Rohrer, D., Taylor, K., Pashler, H., Wixted, J. T., & Cepeda, N. J. (2005). The effect of overlearning on long-term retention. *Applied Cognitive Psychology, 19,* 361–374.

*Rose, R. J. (1984). Processing time for repetitions and the spacing effect. *Canadian Journal of Psychology, 83,* 537–550.

*Rose, R. J. (1992). Degree of learning, interpolated tests, and rate of forgetting. *Memory & Cognition, 20,* 621–632.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86,* 638–641.

*Ross, B. H., & Landauer, T. K. (1978). Memory for at least one of two items: Test and failure of several theories of spacing effects. *Journal of Verbal Learning and Verbal Behavior, 17,* 669–680.

Ruch, T. C. (1928). Factors influencing the relative economy of massed and distributed practice in learning. *Psychological Review, 35,* 19–45.

*Rumelhart, D. E. (1968). The effects of interpresentation intervals on performance in a continuous paired-associate task (Doctoral dissertation, Stanford University, 1967). *Dissertation Abstracts International, 28,* 4782.

*Russo, R., Parkin, A. J., Taylor, S. R., & Wilks, J. (1998). Revising current two-process accounts of spacing effects in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 161–172.

*Schwartz, M. (1975). The effect of constant vs. varied encoding and massed vs. distributed presentations on recall of paired associates. *Memory & Cognition, 3,* 390–394.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261–281). New York: Russell Sage Foundation.

*Shaughnessy, J. J. (1976). Persistence of the spacing effect in free recall under varying incidental learning conditions. *Memory & Cognition, 4,* 369–377.

*Shaughnessy, J. J. (1977). Long-term retention and the spacing effect in free-recall and frequency judgments. *American Journal of Psychology, 90,* 587–598.

*Shaughnessy, J. J., Zimmerman, J., & Underwood, B. J. (1972). Further evidence on the MP-DP effect in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 11,* 1–12.

*Shaughnessy, J. J., Zimmerman, J., & Underwood, B. J. (1974). The spacing effect in the learning of word pairs and the components of word pairs. *Memory & Cognition, 2,* 742–748.

*Shuell, T. J. (1981). Distribution of practice and retroactive inhibition in free-recall learning. *Psychological Record, 31,* 589–598.

*Siegel, M. A., & Misselt, A. L. (1984). Adaptive feedback and review paradigm for computer-based drills. *Journal of Educational Psychology, 76,* 310–317.

*Silverstein, L. D. (1978). Repetition and distribution effects on memory: A psychophysiological analysis (Doctoral dissertation, University of Florida, 1977). *Dissertation Abstracts International, 38,* 5618.

*Simon, J. L. (1979). What do Zielske's real data really show about pulsing? *Journal of Marketing Research, 16,* 415–420.

*Singh, S. N., Mishra, S., Bendapudi, N., & Linville, D. (1994). Enhancing memory of television commercials through message spacing. *Journal of Marketing Research, 31,* 384–392.

*Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology, 30,* 641–656.

*Strong, E. C. (1973). The effects of repetition in advertising: A field experiment (Doctoral dissertation, Stanford University, 1972). *Dissertation Abstracts International, 33,* 4615.

*Strong, E. K., Jr. (1916). The factors affecting a permanent impression developed through repetition. *Journal of Experimental Psychology, 1,* 319–338.

*Stroud, J. B., & Johnson, E. (1941). /1942). The temporal position of reviews. *Journal of Educational Research, 35,* 618–622.

SuperMemo World. (n.d.). *Super memory: Forget about forgetting.* Retrieved March 16, 2006, from http://www.supermemo.com/

Taylor, M. J., & White, K. R. (1992). An evaluation of alternative methods for computing standardized mean difference effect size. *Journal of Experimental Education, 61,* 63–72.

*Thios, S. J. (1972a). The effect of spacing of repetitions and variation of context on memory for discrete events (Doctoral dissertation, University of Virginia, 1971). *Dissertation Abstracts International, 32,* 4909.

*Thios, S. J. (1972b). Memory for words in repeated sentences. *Journal of Verbal Learning and Verbal Behavior, 11,* 789–793.

*Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior, 15,* 529–536.

Thorndike, E. L. (1912). The curve of work. *Psychological Review, 19,* 165–194.

*Toppino, T. C. (1991). The spacing effect in young children's free recall: Support for automatic-process explanations. *Memory & Cognition, 19,* 159–167.

*Toppino, T. C. (1993). The spacing effect in preschool children's free recall of pictures and words. *Bulletin of the Psychonomic Society, 31,* 27–30.

*Toppino, T. C., & DeMesquita, M. (1984). Effects of spacing repetitions on children's memory. *Journal of Experimental Child Psychology, 37,* 637–648.

*Toppino, T. C., & DiGeorge, W. (1984). The spacing effect in free recall emerges with development. *Memory & Cognition, 12,* 118–122.

*Toppino, T. C., & Gracen, T. F. (1985). The lag effect and differential organization theory: Nine failures to replicate. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 185–191.

*Toppino, T. C., Hara, Y., & Hackman, J. (2002). The spacing effect in the

free recall of homogenous lists: Present and accounted for. *Memory & Cognition, 30,* 601–606.

*Toppino, T. C., & Schneider, M. A. (1999). The mix-up regarding mixed and unmixed lists in spacing-effect research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1071–1076.

*Tsai, L.-S. (1927). The relation of retention to the distribution of relearning. *Journal of Experimental Psychology, 10,* 30–39.

*Tzeng, O. J. L. (1973). Stimulus meaningfulness, encoding variability, and the spacing effect. *Journal of Experimental Psychology, 99,* 162–166.

*Underwood, B. J. (1951). Studies of distributed practice: II. Learning and retention of paired-adjective lists with two levels of intra-list similarity. *Journal of Experimental Psychology, 42,* 153–161.

*Underwood, B. J. (1952a). Studies of distributed practice: VI. The influence of rest-interval activity in serial learning. *Journal of Experimental Psychology, 43,* 329–340.

*Underwood, B. J. (1952b). Studies of distributed practice: VII. Learning and retention of serial nonsense lists as a function of intralist similarity. *Journal of Experimental Psychology, 44,* 80–87.

*Underwood, B. J. (1953). Studies of distributed practice: XI. An attempt to resolve conflicting facts on retention of serial nonsense lists. *Journal of Experimental Psychology, 45,* 355–359.

*Underwood, B. J. (1954). Studies of distributed practice: XII. Retention following varying degrees of original learning. *Journal of Experimental Psychology, 47,* 294–300.

Underwood, B. J. (1961). Ten years of massed practice on distributed practice. *Psychological Review, 68,* 229–247.

*Underwood, B. J. (1969). Some correlates of item repetition in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 8,* 83–94.

*Underwood, B. J. (1970). A breakdown of the total-time law in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 9,* 573–580.

Underwood, B. J., & Ekstrand, B. R. (1967). Effect of distributed practice on paired-associate learning. *Journal of Experimental Psychology, 73,* 1–21.

*Underwood, B. J., Kapelak, S. M., & Malmi, R. A. (1976). The spacing effect: Additions to the theoretical and empirical puzzles. *Memory & Cognition, 4,* 391–400.

Underwood, B. J., & Keppel, G. (1963). Retention as a function of degree of learning and letter-sequence interference. *Psychological Monographs: General and Applied, 77* (4, Whole No. 567).

*Underwood, B. J., Keppel, G., & Schulz, R. W. (1962). Studies of distributed practice: XXII. Some conditions which enhance retention. *Journal of Experimental Psychology, 64,* 355–363.

*Underwood, B. J., & Richardson, J. (1955). Studies of distributed practice: XIII. Interlist interference and the retention of serial nonsense lists. *Journal of Experimental Psychology, 50,* 39–46.

*Underwood, B. J., & Richardson, J. (1957). Studies of distributed practice: XVII. Interlist interference and the retention of paired consonant syllables. *Journal of Experimental Psychology, 54,* 274–279.

*von Wright, J. M. (1971). Effects of distributed practice and distributed recall tests on later recall of paired associates. *Journal of Verbal Learning and Verbal Behavior, 10,* 311–315.

*Watts, D., & Chatfield, D. (1976). Response availability as a function of massed and distributed practice and list differentiation. *Psychological Record, 26,* 487–493.

*Waugh, N. C. (1967). Presentation time and free recall. *Journal of Experimental Psychology, 73,* 39–44.

*Waugh, N. C. (1970). On the effective duration of a repeated word. *Journal of Verbal Learning and Verbal Behavior, 9,* 587–595.

*Welborn, E. L. (1933). A study of logical learning in college classes. *20th Annual Conference on Educational Measurement, Indiana University School of Education Bulletin, 10,* 12–20.

*Wells, J. E., & Kirsner, K. (1974). Repetition between and within modalities in free recall. *Bulletin of the Psychonomic Society, 2,* 395–397.

*Wenger, S. K. (1979). The within-list distributed practice effect: More evidence for the inattention hypothesis. *American Journal of Psychology, 92,* 105–113.

Wickelgren, W. A. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology, 9,* 418–455.

*Wilson, J. T. (1949). The formation and retention of remote associations in rote learning. *Journal of Experimental Psychology, 39,* 830–838.

*Wilson, W. P. (1976). Developmental changes in the lag effect: An encoding hypothesis for repeated word recall. *Journal of Experimental Child Psychology, 22,* 113–122.

Wixted, J. T. (2004). On common ground: Jost's (1897) law of forgetting and Ribot's (1881) law of retrograde amnesia. *Psychological Review, 111,* 864–879.

Wixted, J. T., & Ebbesen, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition, 25,* 731–739.

Woźniak, P. A., & Gorzelańczyk, E. J. (1994). Optimization of repetition spacing in the practice of learning. *Acta Neurobiologiae Experimentalis, 54,* 59–62.

*Wright, J., & Brelsford, J. (1978). Changes in the spacing effect with instructional variables in free recall. *American Journal of Psychology, 91,* 631–643.

*Young, J. L. (1966). Effects of intervals between reinforcements and test trials in paired-associate learning (Doctoral dissertation, Stanford University, 1966). *Dissertation Abstracts International, 27,* 3699.

*Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society, 15,* 41–44.

*Zimmerman, J. (1975). Free recall after self-paced study: A test of the attention explanation of the spacing effect. *American Journal of Psychology, 88,* 277–291.

Appendix

Effects of Task Type on Distributed Practice

One lingering concern with our lag analyses is whether task type plays a role in the expression of joint effects between ISI and retention interval. Put another way, is it reasonable to expect the joint effects of ISI difference and retention interval to be constant, regardless of task type? We can think of no a priori reason to expect lag effects to vary on the basis of task type. On the other hand, different experimental methodologies, which vary consistently with task type, might reduce our ability to glean the joint effects of ISI difference and retention interval. Specifically, some paradigms provided consistent and accurate manipulation of ISI difference and retention interval, and these well-controlled paradigms were used in most of the experiments with paired associate tasks. In most experiments with paired associate tasks, items separated by a given lag were almost always followed by exactly the same retention interval. Thus, there is no question that ISI and retention interval values used in this meta-analysis were accurate. In contrast, list recall paradigms did not accurately control ISI difference and retention interval, so there is some degree of incorrectness in the ISI difference and retention interval values we used. To illustrate the problem, say items are represented by i_x. The following is a sample list recall paradigm. Lag is always 1 item, and there are no filler items. The typical primacy and recency buffers have been removed:

$i_1\ i_2\ i_1\ i_2\ i_3\ i_4\ i_3\ i_4\ i_5\ i_6\ i_5\ i_6$

retention interval (time = x)

recall test (unlimited time given to complete test).

The first feature to notice is that retention interval for items i_1 and i_2 is longer than retention interval for i_5 and i_6. This problem becomes worse when list length is long and retention interval is short. Also, we have presented a best-case scenario. Many list recall paradigms present items i_1–i_6, and then rerandomize item order before re-presenting the entire list. This introduces even more variability, as ISI difference is then variable, as is retention interval. An additional, smaller, problem is that giving unlimited time to recall means that retention interval becomes more variable than if recall time were fixed, as occurs in many paired associate paradigms.

(*text continues on page 380*)

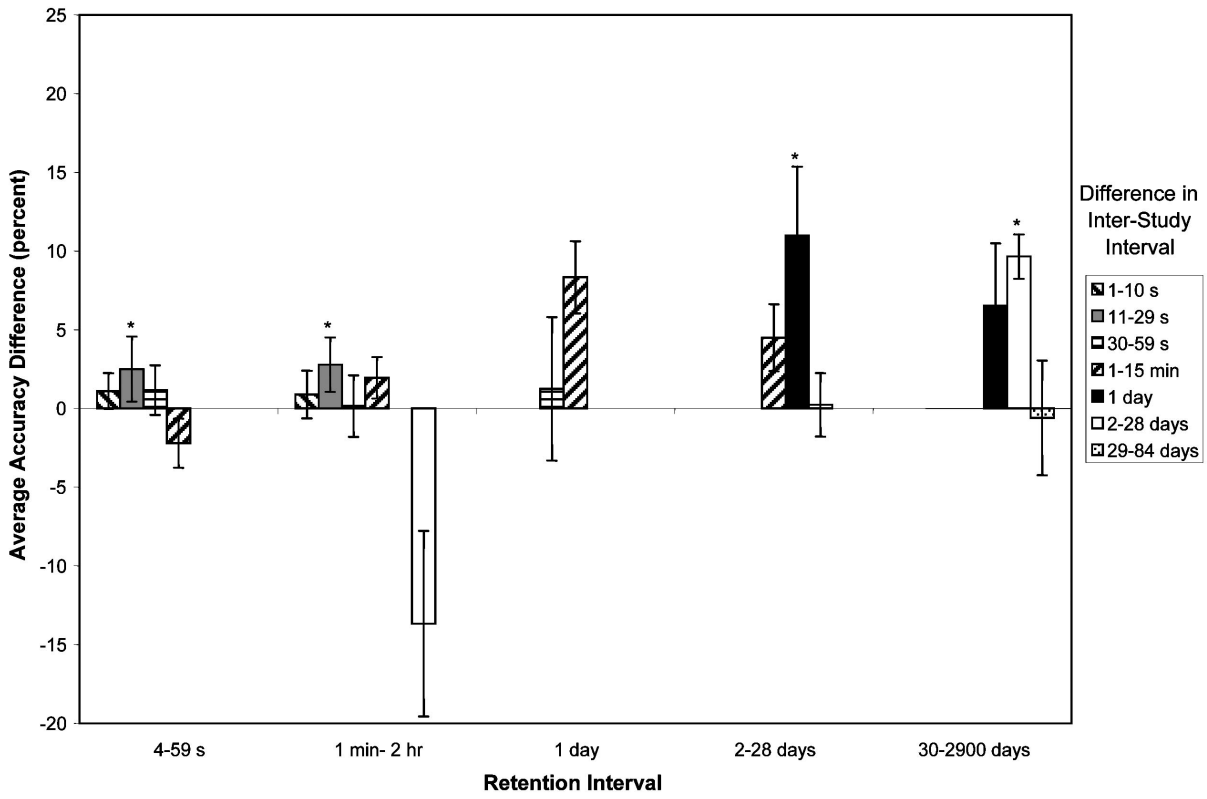

Figure A1. For paired associate studies in the accuracy difference lag analyses, accuracy difference between all adjacent pairs of interstudy interval (ISI) values from each study, binned by difference in ISI and retention interval and averaged across studies. When surrounded by ISI bins with lower accuracy values, the ISI bin showing the highest accuracy value at each retention interval bin is indicated with an asterisk. Error bars represent one standard error of the mean.

(*Appendix continues*)

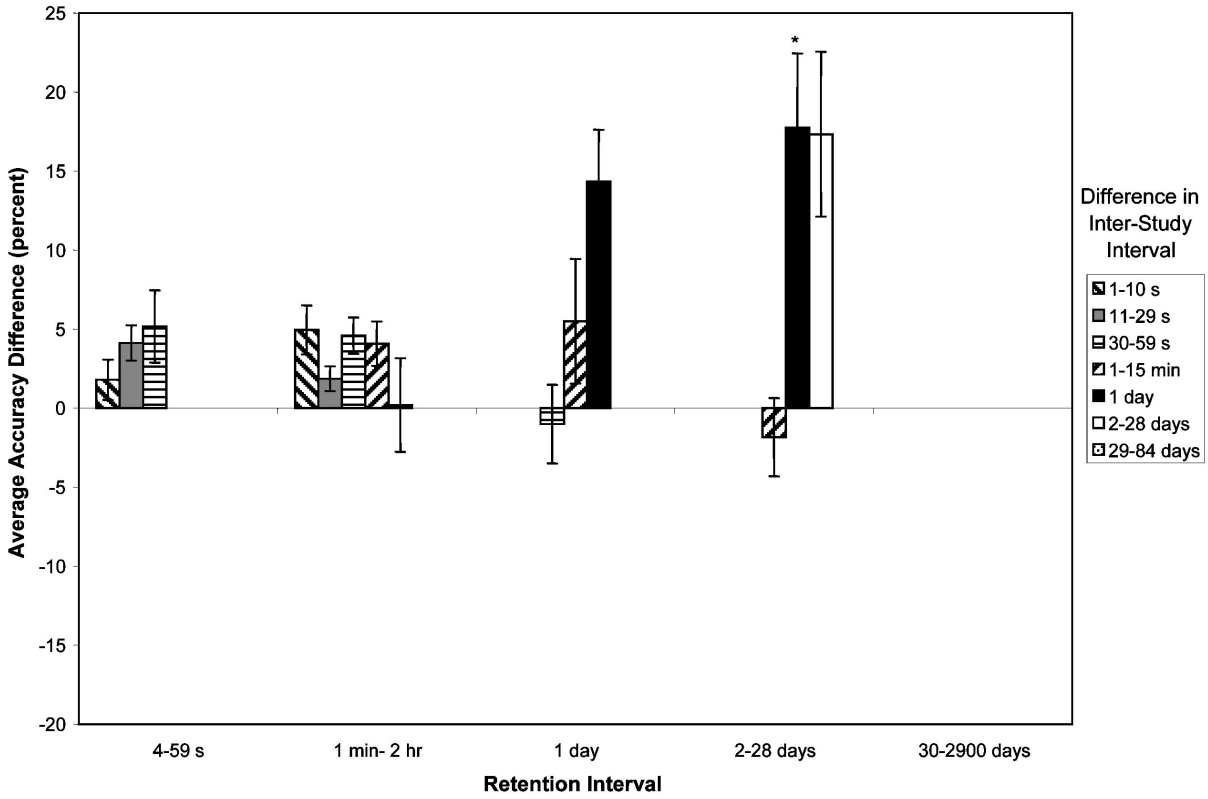

Figure A2. For list recall studies in the accuracy difference lag analyses, accuracy difference between all adjacent pairs of interstudy interval (ISI) values from each study, binned by difference in ISI and retention interval and averaged across studies. When surrounded by ISI bins with lower accuracy values, the ISI bin showing the highest accuracy value at each retention interval bin is indicated with an asterisk. Error bars represent one standard error of the mean.

Table A1
For Paired Associate Data, Shorter and Longer Interstudy Interval (ISI) Range, Retention Interval Range, Percentage Correct at the Shorter and Longer ISI Range, and Statistical Analyses, for Accuracy Difference Lag Analyses

ISI range		Retention interval range	% Correct at ISI range			
Shorter	Longer		Shorter	Longer	*SEM*	Statistical analysis
1–10 s	11–29 s	4–59 s	1.1	2.5	1.6	$t(53) = 0.6$, $p = .522$
11–29 s	1–15 min	4–59 s	2.5	−2.2	1.8	$t(23) = 1.1$, $p = .280$
1–10 s	11–29 s	1 min–2 hr	1.0	2.8	1.5	$t(35) = 0.6$, $p = .554$
11–29 s	2–28 days	1 min–2 hr	2.8	−13.7	3.8	$t(28) = 3.0$, $p < .01$
30–59 s	1–15 min	1 day	1.3	8.3	3.4	$t(8) = 1.5$, $p = .162$
1–15 min	1 day	2–28 days	4.5	11.0	3.3	$t(9) = 1.4$, $p = .194$
1 day	2–28 days	2–28 days	11.0	0.2	2.8	$t(10) = 2.5$, $p < .05$
1 day	2–28 days	30–2,900 days	6.5	9.7	2.7	$t(14) = 1.0$, $p = .356$
2–28 days	29–84 days	30–2,900 days	9.7	−0.6	2.6	$t(16) = 3.2$, $p < .01$

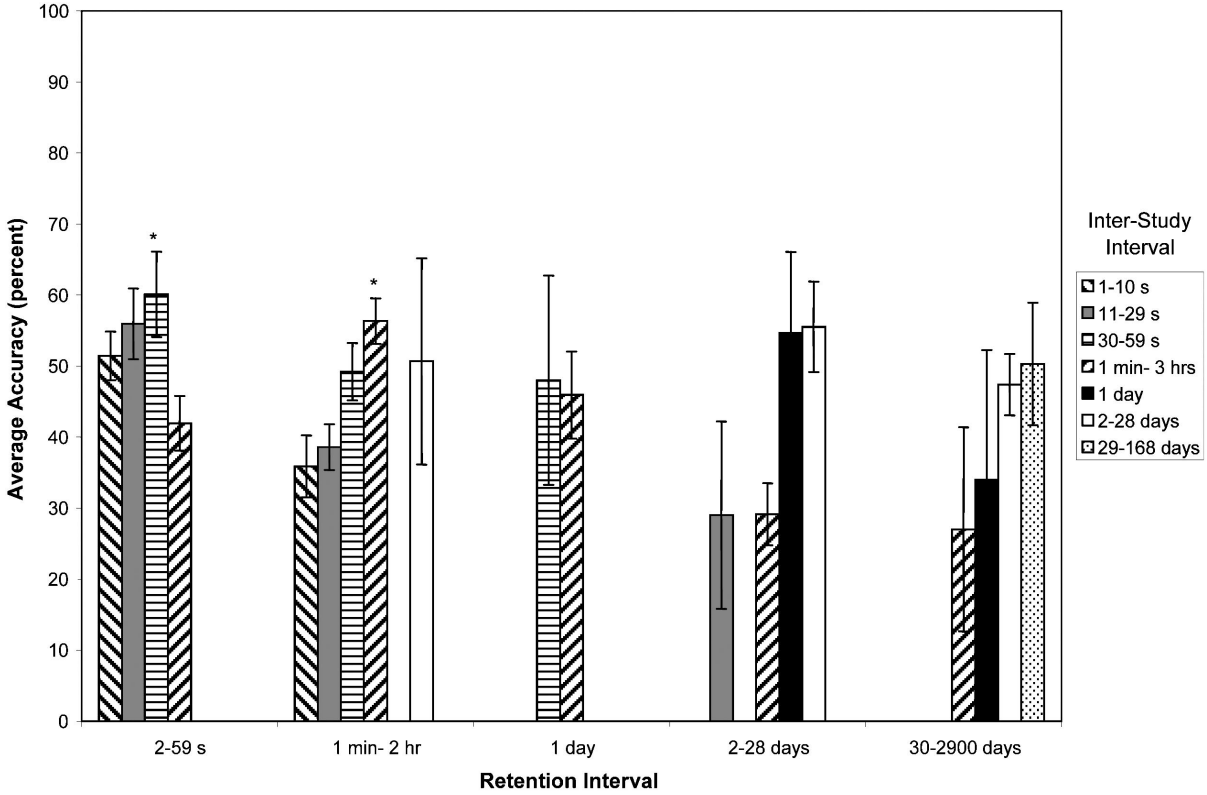

Figure A3. For paired associate studies in the absolute lag analyses, accuracy, binned by interstudy interval (ISI) and retention interval and averaged across studies. When surrounded by ISI bins with lower accuracy values, the ISI bin showing the highest accuracy value at each retention interval bin is indicated with an asterisk. Error bars represent one standard error of the mean.

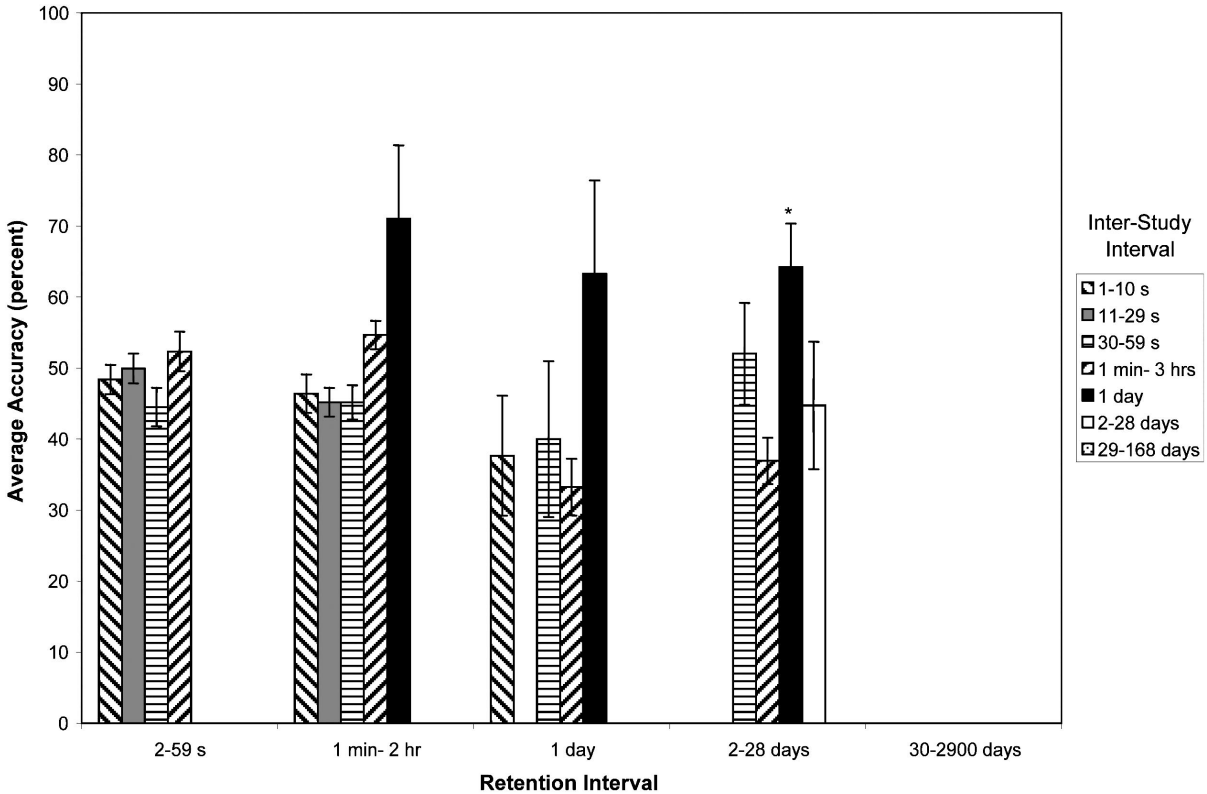

Figure A4. For list recall studies in the absolute lag analyses, accuracy, binned by interstudy interval (ISI) and retention interval and averaged across studies. When surrounded by ISI bins with lower accuracy values, the ISI bin showing the highest accuracy value at each retention interval bin is indicated with an asterisk. Error bars represent one standard error of the mean.

(*Appendix continues*)

Table A2
For Paired Associate Data, Shorter and Longer Interstudy Interval (ISI) Range, Retention Interval Range, Percentage Correct at the Shorter and Longer ISI Range, and Statistical Analyses, for Absolute Lag Analyses

ISI range		Retention interval range	% Correct at ISI Range			Statistical analysis
Shorter	Longer		Shorter	Longer	*SEM*	
1–10 s	30–59 s	2–59 s	51.4	60.1	4.7	$t(60) = 1.3$, $p = .183$
30–59 s	1 min–3 hr	2–59 s	60.1	41.9	4.9	$t(33) = 2.3$, $p < .05$
1–10 s	1 min–3 hr	1 min–2 hr	35.9	56.3	3.8	$t(67) = 3.8$, $p < .001$
1 min–3 hr	2–28 days	1 min–2 hr	56.3	50.7	8.9	$t(46) = 0.4$, $p = .664$
11–29 s	2–28 days	2–28 days	29.0	55.5	9.8	$t(16) = 1.9$, $p = .073$
1 min–3 hr	29–168 days	30–2,900 days	27.0	50.3	11.5	$t(12) = 1.4$, $p = .180$

To assess the impact of these paradigmatic issues, we have reanalyzed lag data, separating by task type. Figures A1 and A2 show joint effects of ISI difference and retention interval, for paired associate and list recall data, respectively. Table A1 provides quantitative analyses of joint effects of ISI difference and retention interval, for paired associate data. As would be predicted by paradigmatic differences, paired associate data paint a much cleaner qualitative picture of joint effects between ISI difference and retention interval. Unfortunately, this cleaner qualitative picture comes with a less clean quantitative picture, because sample size, and thus power, is reduced as well.

In Figures A3 and A4 we present joint effects of absolute ISI and retention interval, for paired associate and list recall data, respectively. Table A2 provides quantitative analyses of joint effects of absolute ISI and retention interval joint effects. The data once again support an increase in optimal ISI as retention interval increases.

Received April 14, 2004
Revision received July 20, 2005
Accepted September 9, 2005 ■