Title
Distributed practice in verbal recall tasks: A review and quantitative synthesis

Permalink
https://escholarship.org/uc/item/3rr6q10c

Journal
Psychological Bulletin, 132(3)

Authors
Nicholas, Cepda
Pashler, Harold
Vul, Ed
et al.

Publication Date
2006

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

Nicholas J. Cepeda
University of Colorado, Boulder

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

Doug Rohrer
University of South Florida

A meta-analysis of the distributed practice effect was performed 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 inter-study interval effects. Analyses suggest that inter-study interval (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, inter-study interval, retention interval

The distributed practice effect refers to an effect of inter-study 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 one second, 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 (one second or longer) separates study episodes for a given item (i.e., 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 and 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
Distributed practice

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 zero items and a time lag of less than one second. 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 to 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 & Radosievich, 1999; Janiszewski, Noel, & Sawyer, 2003; Lee, T. D., & Genovese, 1988; Moss, 1996). The authors of these papers all conclude that distributed practice produces an overall increase in retention, and they argue 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 percent 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’ review concluded that distributed practice has little or no effect on acquisition of motor skills. (Adams claims that most research on distributed practice in motor learning ended in the 1960s due to disinterest, after Hull’s, 1943, learning theory was shown to poorly account for existing data. In concurrence with T. D. Lee and Genovese’s review, Hull’s theory suggested that distributed practice should improve motor learning.)

In their meta-analysis of the distributed practice literature, Donovan and Radosievich (1999) inspected 63 articles that used a wide range of tasks. They examined the effects of several moderators: methodological rigor (on a three 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 hour, 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 Radosievich examined was the interaction of ISI and task domain. Importantly, 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 Radosievich 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, this 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 percent 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 vs. pictorial stimuli, novel vs. familiar stimuli, unimodal vs. bimodal stimulus presentation (e.g., auditory vs. auditory plus visual), structural vs. semantic cue relationships, and isolated vs. 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 non-meaningful stimuli), stimulus
and task. Distributed practice through a complex interplay of time practice benefits are robust, temporal moderators affect 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 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 suggest 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 suggests that increasingly distributed practice impairs learning, seemingly counter to Janiszewski et al.’s (2003) review, which concluded that increasingly distributed practice improved retention. Looking more closely at Donovan and Radosevich’s findings, skill acquisition studies showed decreased final-test learning with longer ISIs, while verbal memory tasks showed non-monotonic 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 suggests 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 paper was 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. Joint effects of ISI and retention interval were assessed 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 where present distributed practice knowledge is severely limited. While the
literature on distributed practice is indeed very large, the present review will disclose (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 utilized 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 the first author using several sources. Lists of potential papers were given to the first author by his coauthors, based on past literature searches for related studies. PsycINFO (1872-2002) and/or ERIC (1966-2002) were searched using 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, destin* rehears*, meta-analysis, 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. For the most analysed analyses, data were separated into relatively small ranges of retention interval (e.g., less than one min, one min-less than 10 min, 10 min-less than one 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 this studies, the reader will not be able to calculate ISI, retention interval, and/or accuracy from the published article.)

Inclusion Criteria

Studies had to meet several criteria to be included. The material was 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 assessed performance at the time of final test. The experiment 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 one 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., one minute equals 0.000694 days and one week equals 7 days). ISI and retention interval were computed based on 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 one min for another item. Because of this confound, we have re-analysed the data, separating out list recall and paired associates studies (see Appendix). For most analyses, data were separated into relatively small ranges of retention interval (e.g., less than one min, one min-less than 10 min, 10 min-less than one 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 this 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 (SD).
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 SD (Morris, 2000; Taylor & White) or various other forms of SD (cf. D’Amico, Neilands, & Zambanaro, 2001; Gleser & Olkin, 1994; Johnson & Eagly, 2000; Shadish & Haddock, 1994) are typically used. In this paper, standard deviation was determined using the method advocated by D’Amico et al., 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 SD were excluded from effect size analyses.

In choosing to use this form of SD, 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 percent. As well, most of the data examined here exhibit neither ceiling nor floor effects, a likely source of unequal variance.

For within-subjects experiments, standard deviation was corrected for dependence between responses using the equation \( SD_{\text{corr}} = SD_{\text{ws}} (2(1 - \rho))^{1/2} \) from Morris and DeShon (2002; cf. Cortina & Nouri, 2000; Dunlap, Cortina, Vaslow, & Burke, 1996; Gibbons, Hedeke, & Davis, 1993), where \( SD_{\text{ws}} \) is the independent groups standard deviation, \( SD_{\text{corr}} \) is the within-subjects standard deviation, and \( \rho \) is the correlation between scores. In the current analysis, correction for dependence used the average of all pair-wise 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 Inter-Study Interval and Retention Interval Joint Effects**

In order 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 comparing adjacent pair-wise within-study experimental conditions. 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, between-condition accuracy differences were computed 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 min. Study 1 also yields two accuracy difference values. For ISIs of 2 days and 1 day, accuracy difference = 50-60 = -10 percent, and for ISIs of 1 day and 1 min, accuracy difference = 60-90 = -30 percent.

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 percent. 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).

Effect sizes were computed 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
Absolute lag analyses. Because we are interested in the relative benefits of specific ISIs, we also performed lag analyses based on 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 percent 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 values for the difference lag analyses.)

To determine the relative benefits of specific ISIs, were are 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, involve between-study comparisons. This is problematic, since overall level of difficulty often differs substantially between studies. Since we have not corrected for these differences, the overall level of difficulty may not be equivalent for every bin. Thus, both absolute and difference analyses are 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- instead of between-study 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: (a) 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 used in the study. The smaller the range of absolute ISIs used, the greater is this potential problem. (b) The actual observed optimal ISI will vary, since 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 vs. 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 one second.) Our analysis of massed vs. 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, since some studies were between-subjects and others were within-subjects.

Spaced presentations led to markedly better final-test performance, compared to massed presentations. For retention intervals less than one minute, spaced presentations improved final-test performance by 9 percent, compared to 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 this massing benefit only occurs with extremely short retention intervals. For example, Peterson, Hillner, and Saltzman (1962) only found a massing benefit when
Distributed Practice

Psychological Bulletin (in press): This manuscript may differ from the final published version
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).

Portions of our data are qualitatively similar to other meta-analysis findings. Like Donovan and Radosevich (1999), our data show non-monotonic effects of ISI difference. Like Janiszewski et al. (2003), our data show generally improved retention as ISI difference increases. Unfortunately, it is impossible to know if we have confirmed these meta-analyses, since 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 inter-study interval lag analyses. Although it is encouraging that difference lag analyses show clear joint effects of ISI difference and retention interval, we are really interested in how absolute ISI interacts with retention interval. Based on the absolute optimal ISI data, we can make concrete recommendations on how large a lag is optimal, given a particular retention interval. Differences in performance between optimal and sub-optimal ISI differences should be smaller and less meaningful as a measure of ideal absolute ISI, compared to differences between optimal and sub-optimal absolute ISIs. This is the case because ISI differences of 7 to 8 days and ISI differences of 0 to 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 one day, and only a few data points used a retention interval longer than one 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 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 3 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, since 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 supra-day, at a single retention interval, are supportive of our findings: Cepeda et al. (2005) presented data in which the optimal ISI was longer than one day at a supra-month retention interval; Gordon (1925) showed that sub-day ISIs are optimal at sub-day retention intervals and that supra-day ISIs are optimal at supra-day 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, & Paulin, 1989; Glenberg, 1976; Peterson, Wampier, 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 one 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 non-monotonic 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.
As noted in the Introduction, in examining commonly used experimental designs, we found that a number of frequently 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 one 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; Strong, E. C., 1973; Strong, E. K., Jr., 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 and E. C. Strong 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 & Tomasetto, 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.

In order 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 non-confounded. There are seven experiments in five papers that used non-confounded designs with ISIs and retention intervals of one 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 confound amount of relearning and ISI, show similar patterns to Cepeda et al., Experiments 2a and 2b, which are unconfounded. The ideal ISI indicated in all these studies is one month or more, at retention intervals of six 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 un-confounded studies by Ausubel, Cepeda et al., Experiment 1, and Glenberg and Lehmann, who used similar retention intervals of 6-10 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 one day. (However, Welborn used a retention interval of 28 days, while Cepeda et al. used a retention interval of 10 days.) Two studies that used unlimited restudy time (Simon, 1979; Strong, E. C., 1973) are in line with similar un-confounded studies (i.e., Ausubel; Cepeda et al., Experiment 1; Glenberg & Lehmann), while one study that used unlimited restudy time (Strong, E. K., Jr., 1916) is not. Even with some inconsistencies between confounded and un-confounded 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 or not, the same basic conclusion would be drawn: optimal ISI increases as retention interval increases.

Expanding vs. Fixed Inter-Study Intervals

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 and 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

*Psychological Bulletin (in press): This manuscript may differ from the final published version*
were used for analyses, as a conservative measure, since some studies were between-subjects ($n = 7$) and others were within-subjects ($n = 11$).

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 one 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 overlap 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 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, since all three of the studies using ISIs and retention intervals longer than one day either provided testing with feedback (Cull) or provided a fixed amount of item restudy time (Clark; Cull; Tsai), which was functionally equivalent to providing feedback (since the entire to-be-learned item was present). With the exception of Cull et al. and Landauer and Bjork (1978), expanding interval studies that used retention intervals less than one day (Cull, 1995; Foos & Smith, 1974; Hser & Wickens, 1989; Siegel & Misselt, 1984) all provided either a fixed amount of restudy time for each item or testing with feedback. We are left with inadequate evidence to support or refute the feedback hypothesis.

General Discussion

While the distributed practice effect has spawned a large literature, prior meta-analyses (Donovan & Radosevich, 1999; T. D. Lee & Genovese, 1988; Janiszewski, 2003) 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 vs. 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 one minute or more than one 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 non-zero value to an even larger non-zero value). When ISI was increased, participants retained more information. However, for long ISIs, in proportion to retention interval, further increases in ISI reduced accuracy. Thus, for a given retention interval, there was a non-zero value of ISI that optimized accuracy. (This is known as a non-monotonic lag effect.) Moreover, the optimal ISI increased as retention interval increased. For instance, at retention intervals of less than one min, ISIs of less than one min maximized retention; at retention intervals of six months or more, ISIs of at least one month maximized retention. These results clearly show that a single ISI does not produce optimal retention across a wide range of retention intervals. The non-monotonic 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 inter-study intervals 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...
interval benefits (Cull et al., 1996; Hser & Wickens, 1989; Landauer & Bjork; Tsai, 1927). We found no obvious systematic differences between studies that do and do not show expanding interval benefits, although one difference that might account for inter-study variability is the presence or absence of feedback. Given the practical import of multi-session 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. While 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 will 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 inter-item associations. (Massed items have associations only to the two immediately adjacent items, while 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 vs. 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.

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’ 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 are 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, 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. While 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 dependence of optimal ISI upon retention interval.

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 non-monotonic 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 Raajimakers, 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 one week, more consolidation into long-term memory will have occurred than if the ISI is one day, and the second trace will inherit this higher state of consolidation. If the delay is too long, say one 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 non-monotonically; 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: by requiring retrieval of the first presentation, a lag effect is found (Thios & D’Agostino); 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). 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 non-monotonic 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 based on our preliminary analysis, study-phase retrieval, consolidation, and encoding variability theories survive as candidate distributed practice theories, while 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 one 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 one 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 one-day interval; learning within a single day impaired learning, compared to a one-day interval between study episodes; learning at one single point in time impaired learning, compared to a several-minute interval between study episodes. The average observed benefit from distributed practice (over massed practice) in these studies was 15 percent, 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 one 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 point out that the file drawer problem tends to be a non-issue 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 inter-study and retention intervals that are educationally relevant, i.e., 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 multi-day 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 non-constant (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 Gorzelaniczky (1994; see also http://www.supermemo.com/) offered a “universal formula” designed to space repetitions at an interval that will produce 95 percent retention, based on Bahrick and Phelps (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 (APA, 2001). As well, it would be useful if researchers reported 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 supra-optimal ISI, the relative merits of expanding (as compared to 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 pair-wise correlations between ISIs, so that dependence between responses can be corrected, whenever the design is within-subjects.

Age effects. Almost all distributed practice data in our analysis (85 percent) are based on performance of young adults (see Table 10). While 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 one day or longer. Until empirical data examining the distributed practice effect in children is 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 practice research; it is hoped that the present review will help researchers to pinpoint where that effort might be the most useful and illuminating.

Author Note

Nicholas J. Cepeda, Department of Psychology, University of California, San Diego, and Department of Psychology, University of Colorado, 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 (US Dept. of Education Grants R305H020061 and R305H040108). Thanks to Jean Trinh for obtaining articles. Thanks to Kelly Braun, Jane Childers, Michael Kahana, and Phil Pavlik for providing raw data. Thanks to Derek Briggs for comments on the manuscript and statistical advice.
References marked with an *asterisk indicate studies included in the meta-analysis.


References

Psychological Bulletin (in press): This manuscript may differ from the final published version


* 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.


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.


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.


Morris, S. B. (2000). Distribution of the standardized mean change effect size for meta-analysis on

* Psychological Bulletin (in press): This manuscript may differ from the final published version


* Psychological Bulletin (in press): This manuscript may differ from the final published version


Appendix

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 based on 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_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 re-randomize item order before re-presenting the entire list. This introduces even more variability, since ISI difference is then variable as well as 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.

To assess the impact of these paradigmatic issues, we have re-analysed 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.
Table 1
Percent Correct on the Final Recall Test for Massed / Spaced Conditions (Average Standard Error of the Mean for Spaced and Massed Trials), Number of Mean Performance Differences / Studies, Total Number of Participants Summing Across All Study / Condition Combinations, and Statistical Analyses, for Spaced Versus Massed Presentations

<table>
<thead>
<tr>
<th>Retention Interval</th>
<th>Percent Correct for Massed / Spaced Conditions</th>
<th>Number of Mean Performance Differences / Studies</th>
<th>Total Number of Participants</th>
<th>Statistical Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-59 s</td>
<td>41.2 / 50.1 (1.7)</td>
<td>105 / 96</td>
<td>5086</td>
<td>t(208)=3.7, p&lt;.001</td>
</tr>
<tr>
<td>1 Min - Less than 10 Min</td>
<td>33.8 / 44.8 (1.5)</td>
<td>124 / 117</td>
<td>6762</td>
<td>t(246)=5.0, p&lt;.001</td>
</tr>
<tr>
<td>10 Min - Less than 1 Day</td>
<td>40.6 / 47.9 (6.1)</td>
<td>11 / 10</td>
<td>870</td>
<td>t(20)=0.6, p=0.535</td>
</tr>
<tr>
<td>1 Day</td>
<td>32.9 / 43.0 (6.0)</td>
<td>15 / 15</td>
<td>1123</td>
<td>t(28)=1.2, p=0.249</td>
</tr>
<tr>
<td>2-7 Days</td>
<td>31.1 / 45.4 (7.3)</td>
<td>9 / 9</td>
<td>435</td>
<td>t(16)=1.4, p=0.190</td>
</tr>
<tr>
<td>8-30 Days</td>
<td>32.8 / 62.2 (8.8)</td>
<td>6 / 6</td>
<td>492</td>
<td>t(10)=2.3, p&lt;0.05</td>
</tr>
<tr>
<td>31 Days or More</td>
<td>17 / 39 (n/a)</td>
<td>1 / 1</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>All Retention Intervals</td>
<td>36.7 / 47.3 (1.1)</td>
<td>271 / 254</td>
<td>14811</td>
<td>t(540)=6.6, p&lt;0.001</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Retention Interval Range</th>
<th>Number of Mean Performance Differences</th>
<th>Number of Data Points</th>
<th>Number of Effect Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-59 s</td>
<td>174</td>
<td>301</td>
<td>14</td>
</tr>
<tr>
<td>1 Min – 2 Hr</td>
<td>259</td>
<td>452</td>
<td>53</td>
</tr>
<tr>
<td>1 Day</td>
<td>27</td>
<td>52</td>
<td>16</td>
</tr>
<tr>
<td>2-28 Days</td>
<td>56</td>
<td>108</td>
<td>31</td>
</tr>
<tr>
<td>30 Days or More</td>
<td>23</td>
<td>34</td>
<td>19</td>
</tr>
</tbody>
</table>

Table 3
Shorter and Longer ISI Range, Retention Interval Range, Percent Correct at the Shorter and Longer ISI Range (Average Standard Error of the Mean), and Statistical Analyses, for Accuracy Difference Lag Analyses

<table>
<thead>
<tr>
<th>Shorter ISI Range</th>
<th>Longer ISI Range</th>
<th>Retention Interval Range</th>
<th>Percent Correct at Shorter / Longer ISI Range</th>
<th>Statistical Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10 s</td>
<td>11-29 s</td>
<td>4-59 s</td>
<td>1.6 / 3.9 (0.9)</td>
<td>t(147)=1.8, p=0.077</td>
</tr>
<tr>
<td>11-29 s</td>
<td>1-15 Min</td>
<td>4-59 s</td>
<td>3.9 / -0.9 (1.2)</td>
<td>t(75)=1.4, p=0.156</td>
</tr>
<tr>
<td>30-59 s</td>
<td>1 Day</td>
<td>1 Min- 2 Hr</td>
<td>3.4 / 1.0 (2.5)</td>
<td>t(61)=0.9, p=0.397</td>
</tr>
<tr>
<td>1-15 Min</td>
<td>1 Day</td>
<td>1 Day</td>
<td>6.4 / 17.5 (2.9)</td>
<td>t(16)=2.7, p&lt;0.05</td>
</tr>
<tr>
<td>1-15 Min</td>
<td>1 Day</td>
<td>2-28 Days</td>
<td>1.5 / 10.3 (2.5)</td>
<td>t(26)=2.4, p&lt;0.05</td>
</tr>
<tr>
<td>1 Day</td>
<td>2-28 Days</td>
<td>2-28 Days</td>
<td>10.3 / 3.5 (2.8)</td>
<td>t(37)=1.7, p=0.091</td>
</tr>
<tr>
<td>1 Day</td>
<td>2-28 Days</td>
<td>30-2900 Days</td>
<td>6.5 / 9.0 (2.7)</td>
<td>t(15)=0.7, p=0.476</td>
</tr>
<tr>
<td>2-28 Days</td>
<td>29-84 Days</td>
<td>30-2900 Days</td>
<td>9.0 / -0.6 (2.6)</td>
<td>t(17)=3.0, p&lt;0.01</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Retention Interval Range</th>
<th>ISI Range</th>
<th>Number of Mean Performance Differences / Studies</th>
<th>Number of Unique Participants</th>
<th>Number of Using Paired Associate</th>
<th>Number of Using List Recall</th>
<th>Number of Using Other Tasks</th>
</tr>
</thead>
</table>

**Psychological Bulletin (in press): This manuscript may differ from the final published version**
### Table 5

**Shorter and Longer ISI Range, Retention Interval Range, Percent Correct at the Shorter and Longer ISI Range (Average Standard Error of the Mean), and Statistical Analyses, for Absolute Lag Analyses**

<table>
<thead>
<tr>
<th>Shorter ISI Range</th>
<th>Longer ISI Range</th>
<th>Retention Interval Range</th>
<th>Percent Correct at Shorter / Longer ISI Range</th>
<th>Statistical Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10 s</td>
<td>30-59 s</td>
<td>2-59 s</td>
<td>49.4 / 54.1 (2.5)</td>
<td>t(162)=1.4, p=.167</td>
</tr>
<tr>
<td>30-59 s</td>
<td>1 Min- 3 Hr</td>
<td>2-59 s</td>
<td>54.1 / 48.9 (2.7)</td>
<td>t(90)=1.3, p=.198</td>
</tr>
<tr>
<td>1-10 s</td>
<td>1 Min- 3 Hr</td>
<td>1 Min- 2 Hr</td>
<td>42.3 / 54.0 (1.7)</td>
<td>t(248)=4.8, p&lt;.001</td>
</tr>
<tr>
<td>1 Min- 3 Hr</td>
<td>2-28 Days</td>
<td>1 Min- 2 Hr</td>
<td>54.0 / 35/7 (4.9)</td>
<td>t(161)=3.4, p&lt;.005</td>
</tr>
<tr>
<td>30-59 s</td>
<td>1 Day</td>
<td>1 Day</td>
<td>36.0 / 62.5 (7.8)</td>
<td>t(16)=2.2, p&lt;.05</td>
</tr>
<tr>
<td>11-29 s</td>
<td>1 Day</td>
<td>2-28 Days</td>
<td>26.4 / 52.8 (7.6)</td>
<td>t(21)=2.5, p&lt;.05</td>
</tr>
<tr>
<td>1 Day</td>
<td>2-28 Days</td>
<td>2-28 Days</td>
<td>52.8 / 45.5 (4.2)</td>
<td>t(58)=1.1, p=.270</td>
</tr>
<tr>
<td>1 Min- 3 Hr</td>
<td>29-168 Days</td>
<td>30-2900 Days</td>
<td>27.0 / 50.3 (11.5)</td>
<td>t(12)=1.4, p=.180</td>
</tr>
</tbody>
</table>

**Note:**
- *ISI*: Interstimulus Interval
- *Tasks*: Number of unique participants using paired associate tasks
- *List Recall*: Number of unique participants using list recall tasks
- *Other Tasks*: Number of unique participants using other tasks
- *Studies*: Number of unique participants using paired associate tasks
- *Tasks*: Number of unique participants using list recall tasks
- *Percent Correct*: Percent correct at the shorter and longer ISI range
- *Statistical Analysis*: t-test results

---

*Psychological Bulletin (in press): This manuscript may differ from the final published version*
Table 6
Number of Data Points / Studies, Number of Unique Participants, and Number of Mean Data Points Using Paired Associate, List Recall, or Other Task Types, for Absolute Lag Analyses, by Retention Interval Range and ISI Range

<table>
<thead>
<tr>
<th>Retention Interval Range</th>
<th>ISI Range</th>
<th>Number of Data Points / Studies</th>
<th>Number of Unique Participants</th>
<th>Number Using Paired Associate Tasks</th>
<th>Number Using List Recall Tasks</th>
<th>Number Using Other Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-59 s</td>
<td>1-10 s</td>
<td>113 / 62</td>
<td>3248</td>
<td>41</td>
<td>66</td>
<td>6</td>
</tr>
<tr>
<td>2-59 s</td>
<td>11-29 s</td>
<td>96 / 57</td>
<td>2694</td>
<td>29</td>
<td>59</td>
<td>8</td>
</tr>
<tr>
<td>2-59 s</td>
<td>30-59 s</td>
<td>51 / 35</td>
<td>1707</td>
<td>21</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>2-59 s</td>
<td>1 Min- 3 Hr</td>
<td>41 / 20</td>
<td>1152</td>
<td>14</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>1 Min- 2 Hr</td>
<td>1-10 s</td>
<td>101 / 66</td>
<td>3711</td>
<td>24</td>
<td>59</td>
<td>18</td>
</tr>
<tr>
<td>1 Min- 2 Hr</td>
<td>11-29 s</td>
<td>84 / 76</td>
<td>4773</td>
<td>25</td>
<td>53</td>
<td>6</td>
</tr>
<tr>
<td>1 Min- 2 Hr</td>
<td>30-59 s</td>
<td>93 / 80</td>
<td>4785</td>
<td>34</td>
<td>40</td>
<td>19</td>
</tr>
<tr>
<td>1 Min- 2 Hr</td>
<td>1 Min- 3 Hr</td>
<td>149 / 83</td>
<td>4867</td>
<td>45</td>
<td>64</td>
<td>40</td>
</tr>
<tr>
<td>1 Min- 2 Hr</td>
<td>1 Day</td>
<td>11 / 8</td>
<td>222</td>
<td>2</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>1 Min- 2 Hr</td>
<td>2-28 Days</td>
<td>14 / 9</td>
<td>390</td>
<td>3</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>1 Day</td>
<td>1-10 s</td>
<td>4 / 4</td>
<td>60</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>1 Day</td>
<td>30-59 s</td>
<td>12 / 11</td>
<td>552</td>
<td>3</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>1 Day</td>
<td>1 Min- 3 Hr</td>
<td>30 / 19</td>
<td>1100</td>
<td>12</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>1 Day</td>
<td>1 Day</td>
<td>6 / 6</td>
<td>83</td>
<td>0</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 7
Final Test Performance for Long-ISI, Long-Retention Interval Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>ISI (Days)</th>
<th>Retention Interval (Days)</th>
<th>Final Test Performance (Percent Correct)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ausubel (1966)</td>
<td>1</td>
<td>6</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>6</td>
<td>40</td>
</tr>
<tr>
<td>Bahrick (1979), Exp 2</td>
<td>1</td>
<td>30</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>30</td>
<td>95</td>
</tr>
<tr>
<td>Bahrick, Bahrick, Bahrick, &amp; Bahrick (1993)</td>
<td>14</td>
<td>360</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>360</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>56</td>
<td>360</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>720</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>720</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>56</td>
<td>720</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>1080</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>1080</td>
<td>62</td>
</tr>
<tr>
<td>Study</td>
<td>ISI (Days)</td>
<td>Retention Interval (Days)</td>
<td>Final Test Performance (Percent Correct)</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>------------</td>
<td>---------------------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>Bahrick &amp; Phelps (1987)</td>
<td>1</td>
<td>2900</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>2900</td>
<td>15</td>
</tr>
<tr>
<td>Burtt &amp; Dobell (1925), Exp 2</td>
<td>3</td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Burtt &amp; Dobell (1925), Exp 3</td>
<td>3</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>17</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td>Cepeda, et al. (2005) Exp 1</td>
<td>1</td>
<td>10</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>10</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>10</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>10</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>10</td>
<td>65</td>
</tr>
<tr>
<td>Cepeda, et al. (2005) Exp 2a</td>
<td>1</td>
<td>168</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>168</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>168</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>84</td>
<td>168</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>168</td>
<td>168</td>
<td>45</td>
</tr>
<tr>
<td>Cepeda, et al. (2005) Exp 2b</td>
<td>1</td>
<td>168</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>168</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>168</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>84</td>
<td>168</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>168</td>
<td>168</td>
<td>17</td>
</tr>
<tr>
<td>Childers &amp; Tomasello (2002), Exp 1</td>
<td>1</td>
<td>1</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>7</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7</td>
<td>53</td>
</tr>
<tr>
<td>Edwards (1917)</td>
<td>1</td>
<td>3</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>4</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>32</td>
</tr>
<tr>
<td>Glenberg &amp; Lehmann (1980), Exp 2</td>
<td>1</td>
<td>7</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>7</td>
<td>25</td>
</tr>
<tr>
<td>Simon (1979)</td>
<td>7</td>
<td>7</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>7</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>35</td>
<td>31</td>
</tr>
<tr>
<td>Spitzer (1939)</td>
<td>1</td>
<td>14</td>
<td>36</td>
</tr>
</tbody>
</table>
Table 8
Percent Correct on the Final Recall Test for Expanding / Fixed Conditions (Average Standard Error of the Mean for Expanding and Fixed Trials), Number of Mean Performance Differences / Studies, Total Number of Participants Summing Across All Study / Condition Combinations, and Statistical Analyses, for Expanding Versus Fixed Study Intervals

<table>
<thead>
<tr>
<th>Retention Interval</th>
<th>Percent Correct for Expanding / Fixed Conditions</th>
<th>Number of Mean Performance Differences / Studies</th>
<th>Total Number of Participants</th>
<th>Statistical Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-59 s</td>
<td>91.0 / 91.0 (n/a)</td>
<td>1 / 1</td>
<td>24</td>
<td>t(18)=0.1, p=.91</td>
</tr>
<tr>
<td>1 Min - Less than 10 Min</td>
<td>49.8 / 48.9 (5.56)</td>
<td>10 / 8</td>
<td>580</td>
<td>t(6)=0.5, p=.65</td>
</tr>
<tr>
<td>10 Min - Less than 1 Day</td>
<td>77.8 / 70.0 (11.5)</td>
<td>4 / 3</td>
<td>614</td>
<td>t(6)=0.5, p=.65</td>
</tr>
<tr>
<td>1 Day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-7 Days</td>
<td>66.3 / 59.5 (10.9)</td>
<td>4 / 3</td>
<td>185</td>
<td>t(6)=0.4, p=.68</td>
</tr>
<tr>
<td>8-30 Days</td>
<td>66.3 / 64.0 (11.2)</td>
<td>3 / 3</td>
<td>115</td>
<td>t(4)=0.1, p=.89</td>
</tr>
<tr>
<td>31 Days or More</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Retention Intervals</td>
<td>62.0 / 58.6 (4.6)</td>
<td>22 / 18</td>
<td>1518</td>
<td>t(42)=0.5, p=.61</td>
</tr>
</tbody>
</table>

Table 9
Percent Correct on Final Test, for Fixed and Expanding Study Intervals, for Studies with a Retention Interval of at Least One Day. (Average) ISI and Retention Interval Are in Days

<table>
<thead>
<tr>
<th>Study</th>
<th>ISI (Days)</th>
<th>Retention Interval (Days)</th>
<th>Fixed Study Intervals (Percent Correct)</th>
<th>Expanding Study Intervals (Percent Correct)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clark (1928)</td>
<td>2</td>
<td>21</td>
<td>63</td>
<td>63</td>
</tr>
<tr>
<td>Cull (2000), Exp 3</td>
<td>2</td>
<td>3</td>
<td>98</td>
<td>84</td>
</tr>
<tr>
<td>Cull (2000), Exp 4</td>
<td>2</td>
<td>8</td>
<td>89</td>
<td>82</td>
</tr>
<tr>
<td>Tsai (1927), Exp 2</td>
<td>2</td>
<td>3</td>
<td>48</td>
<td>61</td>
</tr>
<tr>
<td>Tsai (1927), Exp 2</td>
<td>2</td>
<td>7</td>
<td>36</td>
<td>46</td>
</tr>
<tr>
<td>Tsai (1927), Exp 3</td>
<td>2</td>
<td>3</td>
<td>56</td>
<td>74</td>
</tr>
<tr>
<td>Tsai (1927), Exp 3</td>
<td>2</td>
<td>17</td>
<td>40</td>
<td>54</td>
</tr>
</tbody>
</table>

Table 10
Number of Mean Performance Differences and Effect Sizes, by Age Group. Ages Are in Years

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Number of Mean Performance Differences</th>
<th>Number of Effect Sizes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preschool</td>
<td>29</td>
<td>13</td>
</tr>
<tr>
<td>Elementary School</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>Junior High</td>
<td>29</td>
<td>3</td>
</tr>
<tr>
<td>High School</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>Young Adult (18-35)</td>
<td>714</td>
<td>97</td>
</tr>
<tr>
<td>Middle-Aged Adult (36-60)</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Older Adult (61+)</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>Mixed Adult (18+)</td>
<td>14</td>
<td>9</td>
</tr>
</tbody>
</table>
Table A1
For Paired Associates Data, Shorter and Longer ISI Range, Retention Interval Range, Percent Correct at the Shorter and Longer ISI Range (Average Standard Error of the Mean), and Statistical Analyses, for Accuracy Difference Lag Analyses

<table>
<thead>
<tr>
<th>Shorter ISI Range</th>
<th>Longer ISI Range</th>
<th>Retention Interval Range</th>
<th>Percent Correct at Shorter / Longer ISI Range</th>
<th>Statistical Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10 s</td>
<td>11-29 s</td>
<td>4-59 s</td>
<td>1.1 / 2.5 (1.6)</td>
<td>t(53)=0.6, p=.522</td>
</tr>
<tr>
<td>11-29 s</td>
<td>1-15 Min</td>
<td>4-59 s</td>
<td>2.5 / -2.2 (1.8)</td>
<td>t(23)=1.1, p=.280</td>
</tr>
<tr>
<td>1-10 s</td>
<td>11-29 s</td>
<td>1 Min- 2 Hr</td>
<td>1.0 / 2.8 (1.5)</td>
<td>t(35)=0.6, p=.554</td>
</tr>
<tr>
<td>11-29 s</td>
<td>2-28 Days</td>
<td>1 Min- 2 Hr</td>
<td>2.8 / -13.7 (3.8)</td>
<td>t(28)=3.0, p&lt;.01</td>
</tr>
<tr>
<td>30-59 s</td>
<td>1-15 Min</td>
<td>1 Day</td>
<td>1.3 / 8.3 (3.4)</td>
<td>t(8)=1.5, p=.162</td>
</tr>
<tr>
<td>1-15 Min</td>
<td>1 Day</td>
<td>2-28 Days</td>
<td>4.5 / 11.0 (3.3)</td>
<td>t(9)=1.4, p=.194</td>
</tr>
<tr>
<td>1 Day</td>
<td>2-28 Days</td>
<td>2-28 Days</td>
<td>11.0 / 0.2 (2.8)</td>
<td>t(10)=2.5, p&lt;.05</td>
</tr>
<tr>
<td>1 Day</td>
<td>2-28 Days</td>
<td>30-2900 Days</td>
<td>6.5 / 9.7 (2.7)</td>
<td>t(14)=1.0, p=.356</td>
</tr>
<tr>
<td>2-28 Days</td>
<td>29-84 Days</td>
<td>30-2900 Days</td>
<td>9.7 / -0.6 (2.6)</td>
<td>t(16)=3.2, p&lt;.01</td>
</tr>
</tbody>
</table>

Table A2
For Paired Associates Data, Shorter and Longer ISI Range, Retention Interval Range, Percent Correct at the Shorter and Longer ISI Range (Average Standard Error of the Mean), and Statistical Analyses, for Absolute Lag Analyses

<table>
<thead>
<tr>
<th>Shorter ISI Range</th>
<th>Longer ISI Range</th>
<th>Retention Interval Range</th>
<th>Percent Correct at Shorter / Longer ISI Range</th>
<th>Statistical Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10 s</td>
<td>30-59 s</td>
<td>2-59 s</td>
<td>51.4 / 60.1 (4.7)</td>
<td>t(60)=1.3, p=.183</td>
</tr>
<tr>
<td>30-59 s</td>
<td>1 Min- 3 Hr</td>
<td>2-59 s</td>
<td>60.1 / 41.9 (4.9)</td>
<td>t(33)=2.3, p&lt;.05</td>
</tr>
<tr>
<td>1-10 s</td>
<td>1 Min- 3 Hr</td>
<td>1 Min- 2 Hr</td>
<td>35.9 / 56.3 (3.8)</td>
<td>t(67)=3.8, p&lt;.001</td>
</tr>
<tr>
<td>1 Min- 3 Hr</td>
<td>2-28 Days</td>
<td>1 Min- 2 Hr</td>
<td>56.3 / 50.7 (8.9)</td>
<td>t(46)=0.4, p=.664</td>
</tr>
<tr>
<td>11-29 s</td>
<td>2-28 Days</td>
<td>2-28 Days</td>
<td>29.0 / 55.5 (9.8)</td>
<td>t(16)=1.9, p=.073</td>
</tr>
<tr>
<td>1 Min- 3 Hr</td>
<td>29-168 Days</td>
<td>30-2900 Days</td>
<td>27.0 / 50.3 (11.5)</td>
<td>t(12)=1.4, p=.180</td>
</tr>
</tbody>
</table>

Figure Captions

Figure 1. Graphical representation of two hypothetical studies, and the difference and absolute lag graphs that would result when performing lag analyses of these studies.

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

Figure 3. For all studies in the accuracy difference lag analyses, accuracy difference between all adjacent pairs of 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 starred. Error bars represent one standard error of the mean.

Figure 4. For all studies in the effect size lag analyses, effect sizes for all adjacent pairs of 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 starred. Error bars represent one standard error of the mean.

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

Figure 6. For all studies in the absolute lag analyses, accuracy, binned by 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 starred. Error bars represent one standard error of the mean.
Figures A1. For paired associate studies in the accuracy difference lag analyses, accuracy difference between all adjacent pairs of 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 starred. Error bars represent one standard error of the mean.

Figures A2. For list recall studies in the accuracy difference lag analyses, accuracy difference between all adjacent pairs of 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 starred. Error bars represent one standard error of the mean.

Figures A3. For paired associate studies in the absolute lag analyses, accuracy, binned by 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 starred. Error bars represent one standard error of the mean.

Figures A4. For list recall studies in the absolute lag analyses, accuracy, binned by 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 starred. Error bars represent one standard error of the mean.
Figure 1

Figure 2

*Psychological Bulletin (in press): This manuscript may differ from the final published version*
Psychological Bulletin (in press): This manuscript may differ from the final published version
Figure 5

Figure 6

*Psychological Bulletin (in press): This manuscript may differ from the final published version*
Figure A1

Figure A2

*Psychological Bulletin (in press): This manuscript may differ from the final published version*
Figure A3

Figure A4

_Psychological Bulletin (in press): This manuscript may differ from the final published version_