Journal of Experimental Psychology: Learning, Memory, and Cognition

Effects of Spaced Versus Massed Training in Function Learning

Mark A. McDaniel, Cynthia L. Fadler, and Harold Pashler

Online First Publication, April 8, 2013. doi: 10.1037/a0032184

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/13/$12.00 DOI: 10.1037/a0032184

Effects of Spaced Versus Massed Training in Function Learning

Mark A. McDaniel
Washington University and Center for Integrative Research on Cognition, Learning, and Education, St. Louis, Missouri

Cynthia L. Fadler
Washington University

Harold Pashler
University of California, San Diego

A robust finding in the literature is that spacing material leads to better retention than massing; however, the benefit of spacing for concept learning is less clear. When items are massed, it may help the learner to discover the relationship between instances, leading to better abstraction of the underlying concept. Two experiments addressed this question through a typical function learning task in which subjects were trained via presentations of input points (cue values) for which output responses (criterion values) were required. Subjects were trained either using spaced points, strategically massed points (points were paired in training such that they occurred on the same side of the underlying V-shaped function), or randomly massed points (points were randomly paired during training). All subjects were then tested on repeated training points, new (interpolation) points within the training range, and extrapolation points that fell outside the training range. Spacing led to superior interpolation and extrapolation performance, with random massing leading to the worst performance on all test trial types. These results suggest that, at least for function concepts, massed training is not superior to spaced training for concept learning.

Keywords: spaced versus massed training, function learning, concept learning

A well-established finding in the literature is that spacing repetitions of target items produces better memory for those targets than does massing their repetition (termed the *spacing effect*; Jacoby, 1978; see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006, for a review). This pattern is robust across materials, obtaining with words presented in a list (Foos & Smith, 1974), pictures in a list (Rea & Modigliani, 1987), related pairs of words in a list (Jacoby), and vocabulary—meaning pairs presented in a list (Dempster, 1987). Spacing effects have also been reported for learning particular mathematical procedures, such as calculating permutations (Rohrer & Taylor, 2007, Experiment 1), and for learning arithmetic skills (Rickard, Lau, & Pashler, 2008). Indeed, the experimental demonstrations of the benefits of spacing (relative to massing) for retention are so compelling that basic memory researchers advocate spaced instruction for improving educational outcomes (see, e.g., Bjork, 1994; Cepeda et al., 2006; Dempster, 1989; Rohrer & Pashler, 2010). Yet, uncertainty remains regarding the relative advantages of spacing compared to massing for a fundamental and educationally important aspect of learning, that of concept learning.

Before developing the present issue regarding the outcome of concept learning with spaced versus massed presentation, it is important to specify that the focus here is on repeated presentations of training stimuli in a spaced or massed fashion within a single training session, paralleling much of the contemporary verbal memory literature (e.g., Jacoby, 1978). Spacing has also referred to the distribution of multiple training sessions over time, with this implementation of spacing typically investigated and applied to studies of skill development and learning of complex domains (e.g., medical training: Moulton et al., 2006; statistics lessons: Smith & Rothkopf, 1984; middle school biology: Reynolds & Glaser, 1964). In this article, we focus on training of a particular conceptual entity within one training session using repeated training stimuli.

In concept learning, learners are exposed to a set of instances that reflect a concept and, through repeated presentation of the instances and feedback on their responses, may come to abstract the essential features and their relations (e.g., Bourne, 1974). A prominent conjecture has been that massing the instances from a particular concept is preferable to spacing the instances of a particular concept across the stimulus set (i.e., across a set that reflects several different categorical responses), because massing allows the learner to more easily extract the similarities that

Mark A. McDaniel, Department of Psychology, Washington University, and Center for Integrative Research on Cognition, Learning, and Education, St. Louis, Missouri; Cynthia L. Fadler, Department of Psychology, Washington University; Harold Pashler, Department of Psychology, University of California, San Diego.

This research was supported by Grant 220020166, Applying Cognitive Psychology to Enhance Educational Practice, from the James S. McDonnell Foundation. The third author's contributions were supported by a Multidisciplinary University Research Initiative award from the Office of Naval Research (25684A) and by Sciences Learning Center Grant SBE-0542013 from the National Science Foundation. We thank Chris Kudelka for assisting with the programming of the training conditions.

Correspondence concerning this article should be addressed to Mark A. McDaniel, Campus Box 1125, Department of Psychology, Washington University, St. Louis, MO 63130. E-mail: mmcdanie@artsci.wustl.edu

characterize the concept (i.e., the common category response; see Kornell & Bjork, 2008). Note that in some paradigms, a particular stimulus can be repeated; in other paradigms, different stimuli from the same category are either massed or spaced across stimuli from other categories, but the same stimulus is never repeated (Kornell & Bjork, 2008). A particularly galvanizing claim is that "spacing is the friend of recall, but the enemy of induction" (Rothkopf, cited in Kornell & Bjork, 2008, p. 585).

Until recently, little experimental work has evaluated this possibility that spacing is inferior to massing for promoting concept learning. Kornell and Bjork (2008) reported that learning of the artistic style of a set of artists was superior when the learning procedure presented interleaved examples of the work of different artists relative to a procedure that massed the examples from each artist. Kornell and Bjork interpreted this finding as demonstrating that spacing was more effective than massing for concept learning. However, the experimental implementation of the spaced-learning condition confounded spacing with interleaving of the instances (from different styles) with one another, thereby clouding the interpretation of these results. In the few studies that have isolated interleaving from spacing, interleaving is found to promote learning by improving the ability of the learners to select the appropriate response for a given instance. Specifically, Rohrer and Taylor (2007) found that in training students to compute the volumes of four different geometric solids, interleaved practice on the four solids fostered much more accurate selection of the appropriate formula than did massed practice; errors in the formulae per se were not reported to be significantly different across interpolated and massed training. This pattern parallels Kornell and Bjork's finding in which learners more accurately identified the correct artist for test paintings, suggesting that the interpolated nature of the "spaced" condition may have played a prominent role in their result (see Kang & Pashler, 2012, for evidence directly supporting this possibility). Thus, the supposition that spacing is the enemy of concept learning remains an open issue.

The issue regarding the confounding of interleaving with spacing notwithstanding, the limited published work on the effects of spacing on concept learning has examined a narrow range of concept learning tasks. As far as we are aware, only a category learning task, and in particular that of artistic style, has been examined (Kang & Pashler, 2012; Kornell & Bjork, 2008). In category learning, inputs are mapped to categorical labels, but other kinds of conceptual structures are possible (Busemeyer, Byun, Delosh, & McDaniel, 1997). Another common structure is one in which continuous input values are mapped to continuous output values—that is, a function. The purpose of the current experiments was to examine the effects of spacing relative to massing in the learning of a function.

We focused on function learning for several reasons. First, the acquisition and use of functional relationships is ubiquitous. For example, interest rates are forecast based on their functional relation with inflation rates, farmers learn to predict harvest yields on the basis of the amount of rainfall, and employers might learn to anticipate how job performance is related to intelligence or other traits (McDaniel & Busemeyer, 2005). Second, and more importantly, a typical function-learning paradigm dovetails with the classic verbal learning paradigms in which spacing has been studied. In both the function learning and the verbal learning paradigms, particular items are repeated in the study session (for function learning the item is a specific input or cue value—e.g., DeLosh, Busemeyer, & McDaniel, 1997; McDaniel, Dimperio, Griego, & Busemeyer, 2009—and the inputs are repeated to support reasonable learning levels). Accordingly, in the function learning paradigm, spacing can be implemented as it is in the verbal learning paradigms, with the spacing of a particular item achieved by inserting other training items between the spaced repetitions. Two aspects of this particular design are worth noting. One, it differs from the interleaving paradigms in which the "spaced" instances are interleaved with items from other categories. In the function learning paradigm, all of the items are generated from the same function, and the subjects' task is to learn that function. Second, following many spacing studies in the memory literature, spacing is instantiated both (a) temporally, in which the repetitions of the training item are separated from each other by a temporal delay, and (b) distributionally, in which the repetitions are separated from each other by presentations of other training items (Dempster, 1989; but see Rea & Modigliani, 1987, for a study in which spacing was only temporal).

A third attractive feature of function learning is that unlike the verbal learning paradigms, the relative benefit of spacing to massing can be assessed at two levels. One is the accuracy of retrieval for the specific input–output points presented in training (i.e., memory for the input–output pairs). The other is the learning of the more general relationship between input and output points (i.e., learning of the function rule) as assessed in transfer to new inputs. As we point out below, it is conceivable that the commonly observed advantage of spacing for learning particular training stimuli (in this case input–output points) might be reversed in the context of function learning. By examining markers of both levels of learning, we hoped to better characterize the outcomes of spacing versus massing on learning. Before developing the predictions, we briefly overview the experiment.

In common with other recent work examining function learning, we selected a V-shaped function for training (Kang, McDaniel, & Pashler, 2011; McDaniel, Cahill, Robbins, & Wiener, in press; McDaniel et al., 2009, Experiment 1). Subjects were trained on 20 training points from within a restricted range of input values (80–120) that encompassed the vertex (input value = 100) of the "V." Training points were repeated 10 times. In the spaced condition, the set of 20 training points was presented once, followed by another presentation of the "list," and so on for 10 blocks (in which each training point was presented only once in each block). For a massed condition, we considered several possible ways to implement massed training. One straightforward procedure that would directly parallel the verbal learning paradigms would have been to repeat each training point 10 times in succession. However, our pilot work indicated that massing in this fashion produced virtually no learning as indexed by a retention test on trained points.

Accordingly, we modified the "standard" massing procedure to try to better align the massing procedure with the theoretical assumption that the advantage of massed training is that it allows the learner to more easily discern the concept (cf. Kornell & Bjork, 2008). Consistent with theoretical assumptions about learning in a multiple-cue prediction task (Klayman, 1988; the inputs are multiple discrete-valued cues and outputs are continuous values), we reasoned that for a function concept, learners must notice and understand how the output (criterion value) changes in relation to the changes in the input (cue value). Consequently, we massed presentation of pairs of training stimuli such that each stimulus was presented followed by a second stimulus, with this pair of presentations repeated five times in succession. Then another pair of stimuli was presented in identical fashion and so on until 10 pairs of stimuli (20 different stimuli) had been presented. This entire sequence was then repeated for a total of 200 training trials.

To provide a more comprehensive comparison between the learning outcomes with spaced versus massed presentations, we examined two instantiations of the just-described massed condition. In the *massed random condition* (used in Experiment 1 only), a random presentation order for the stimuli was used, and the first two stimuli from the random order were paired, the third and fourth stimuli were paired, and so on. As would be expected, the random order produced some pairings that contained stimuli on both sides of the function's vertex and these pairings would not make apparent the different slopes on each side of the vertex, thereby potentially exacerbating the difficulty of abstracting a non-linear (quadratic-like) function (cf. Pachur & Olsson, 2012, Experiment 3).

Accordingly, in another massed condition, all massed pairs of stimuli were sampled from one or the other side of the function (labeled the *massed strategic* condition; used in Experiments 1 and 2). In this condition, massed pairings would clearly convey the relation between changes in input values (x) and changes in output values (y) for each side of the function, and therefore it seemed possible that this massed presentation condition would well convey the function rule. After training, learners were tested on some of the original training stimuli, new stimuli within the training range (interpolation), and new stimuli outside the training range (extrapolation).

One straightforward prediction based on the spacing effect with verbal materials is that for trained stimuli, learners in the spaced training condition should more accurately remember the output (criterion value) associated with each input (cue value) than learners in the massed condition. The more interesting question pertains to the transfer test, particularly those probes that required extrapolating beyond the training examples. In this regard, it is important to note that learning of the trained input–output pairs does not necessarily implicate learning of the abstract relation among the pairs. For purposes of exposition, we label this abstract component the *function rule* (in adopting this label we do not intend to imply any particular formal rule representation; cf. McDaniel & Busemeyer, 2005). An associative-learning model that does not acquire relational information (the function rule) approximates human learning performance in these tasks (Busemeyer et al., 1997), and in line with this model, some human subjects display highly accurate learning of the training stimuli with little or no apparent learning of the function (as indicated by flat extrapolation; McDaniel et al., in press; (The associative model includes some generalization in activation of input values and output values, thereby supporting accurate interpolation; Busemeyer et al., 1997.) Consequently, it is possible that although spacing might promote accurate retention of the training stimuli, it will not stimulate learning of the functional relation among the stimuli for a significant portion of learners.

By contrast, massing pairs of training stimuli might increase the salience of the relations among training stimuli, especially specific information about the magnitude of the change in the output associated with particular changes in the input values, thereby assisting learning of the functional relation (Klayman, 1988; Pachur & Olsson, 2012). This possibility is suggested by a recent finding reported in a multiple-cue prediction task (Pachur & Olsson, 2012). Learners who were presented with pairs of stimuli and encouraged to focus on comparisons among these paired training stimuli tended to learn the abstract relation (the multiple-cue function) between cue combinations and criterion values, whereas learners who focused on predicting the criterion value for each individual multiple-cue stimulus (i.e., a standard spaced presentation) were more likely to just learn the associations between the particular inputs (cue combinations) and outputs (criterion value).

Similarly, in the current study, massing presentation of stimulus pairs may encourage learners to attend to the relation between the input–out pairs (change in outputs with the change in inputs). Even though each item of the pair is presented sequentially, not concurrently as in Pachur and Olsson's (2012) paradigm, one would expect the massed presentations of pairs of stimuli to reduce demands on predicting criterion values, because these stimuli can be easily maintained in working memory. Reducing demands on predicting individual criterion values may be one parameter that allows learners to instead focus attention on relations among the stimuli. For several reasons then, massing presentation of stimulus pairs may more likely be congruent with *comparative hypothesizing* (Klayman, 1988) processes essential for learning the function relation between inputs and outputs than would spacing presentation.

On the above account, most if not all learners given massed training will gain some sense of the general shape of the function, whereas a fair proportion of learners given spaced training will not learn much about the function shape. This leads to the prediction that extrapolation performance should be better after massed training than after spaced training. To evaluate this prediction, we examined two indices of extrapolation: the mean absolute error of the learner's predicted output relative to the given function value and the slope of the learner's extrapolation predictions. A slope of zero suggests that learners have learned little if anything about how output values change as a function of the change in input values (the function rule). Note that a learner may not remember exactly the output paired with a particular input, but if the learner has learned something about the shape of the function relating the trained inputs and outputs then the slopes displayed in extrapolation would still resemble the slopes reflected by the training items. Therefore, if the massed training leads to better acquisition of the general function shape, then the slopes in extrapolation should be steeper (more different from zero) after massed training than after spaced training.

It is worth mentioning that for the massed training condition, the posited reliance on working memory for preparing output responses to individual input values would be expected to result in fragile long-term memory representations of the individual training stimuli (cf. Craik & Watkins, 1973; Jacoby, 1978). Consequently, learners in the massed condition could display relatively poor accuracy in the test phase for the trained stimuli, even though they produce highly accurate responding during training and a sense of the function shape (as evidenced on extrapolation test trials).

Alternatively, a common view is that during massed presentations, learners become inattentive to the items, thereby reducing the quality or quantity of processing, or both (Hintzman, 1974). Reduced attention to the presentations following the first presentation may not only penalize memory for the trained input–output pairs (thereby extending the spacing-effect patterns with verbal materials) but could also diminish learners attending to the changes in outputs as a function of changes in inputs on some trials, thereby hampering induction of the functional rule. From this perspective, acquisition of the function shape might be expected to be no better or worse after massed training than spaced training.

Experiment 1

Method

Subjects and design. Sixty undergraduates from the subject pool at Washington University in St. Louis participated for partial fulfillment of a course requirement or \$5 per half hour of participation. Subjects were randomly assigned to one of three conditions (*Spaced*, *Massed Random*, *Massed Strategic*) with 20 subjects in each condition. The final test consisted of three trial types (*Trained*, *Interpolation*, *Extrapolation*) and was identical for all participants, resulting in a mixed 3 (condition: *Spaced*, *Massed Random*, *Massed Strategic*) $\times$ 3 (trial type: *Trained*, *Interpolation*, *Extrapolation*) factorial design.

Stimuli. As described above, subjects were trained on a set of input–output pairs that conformed to a continuous bilinear function. The function itself was V-shaped within the interval $51 \leq x \leq 149$ with the vertex at 100. For values of $x \leq 100$, $f(x) = 230 - 2.2x$, and for values of $x \geq 100$, $f(x) = 2.2x - 210$, including only integers for both inputs and outputs. During training, subjects were presented with 20 distinct trials (odd integers between 80 and 120) 10 times each, for a total of 200 trials during training. For individuals in the spaced condition, the 20 training trials were randomized within each of 10 blocks, and all subjects received a single order of presentation. For each of the massed conditions, participants received a pair of inputs 5 times each for a total of 10 trials before moving onto the next pair of trials. After all 10 pairs had been trained (100 trials), participants received the same trials again for a total of 200 trials.[1] The pairs in the massed random condition were randomly chosen, with a different random order constructed for each participant for the first 100 trials (an example of the trials for two sets of massed input pairs is as follows: 103, 111, 103, 111, 103, 111, 103, 111, 103, 111, 113, 89, 113, 89, 113, 89, 113, 89, 113, 89). The ordering of pairings for the first 100 trials was repeated in the same order for the second 100 trials (keeping pairs intact). The pairs in the massed strategic condition were chosen by randomly ordering items on each side of the vertex, with all participants given a single random order of pairs (e.g., 101, 119, 101, 119, 101, 119, 101, 119, 101, 119, 97, 87, 97, 87, 97, 87, 97, 87, 97, 87); the same order was then repeated in the second 100 training trials. The procedure for the massed strategic condition was designed to give subjects the best opportunity to learn the function, as pairs on a single side would be related with a single slope, whereas items on opposite sides of the vertex (possible in the massed random condition) would be governed by two different slopes.

During the final test, subjects were presented with a total of 60 trials; 20 of these were the odd integers presented during training (*Trained* trials), 20 were novel even integers within the trained interval (*Interpolation* trials), and the final 20 were outside the training range (*Extrapolation* trials; nine randomly chosen odd integers between 51 and 79, and 11 randomly chosen odd integers between 121 and 149). Subjects were presented all 60 trials in a single random order.

Procedure. Subjects were tested in small groups of 1–4 individuals. All instructions were self-paced and presented to subjects on the computer screen. Subjects were told to imagine that they had been hired by the National Aeronautics and Space Administration (NASA) to investigate an organism that has been discovered on Mars. The organism absorbs a newly discovered element, Zebon, and then releases another element, Beros. Subjects were then told to try to determine how much Beros will be released given different amounts of Zebon. Subjects were given a sample trial to ensure that they understood how to make a specific predicted value with the arrow keys. On each trial, subjects were shown an input value bar (“Zebon Absorbed”), a bar showing their predicted value (“Your Prediction”; always initially positioned at zero), and a bar that displayed the correct answer as feedback (“Beros Released”). Subjects were given unlimited time to respond by adjusting the height of the bar using the arrow keys, and, after pressing enter to submit their prediction, subjects were given immediate feedback. Feedback consisted of the output value displayed on the “Beros Released” bar and a sentence stating, “Your prediction was __ units off.” In addition, depending on the accuracy of the prediction, subjects saw the following feedback: “Perfect!”; “Great job!”; “Good job.”; or “Not bad.” This feedback appeared on the screen for 4 s to ensure that all participants received the same amount of study time before the next trial automatically appeared.

Immediately following training, participants began the test. They were given the same instructions again but were told that they would not receive feedback. During the test, subjects were presented with 60 trials in a single random order. Each trial consisted of only two bars, the input bar (“Zebon Absorbed”) and the predicted response (“Your Prediction”; always initially positioned at zero). Subjects again adjusted the height of the response bar by using the arrow keys and pressed enter to submit their response. The trials were again self-paced, but no feedback was provided. After each response, subjects were told to wait for the next trial for 5 s, and then the program proceeded to the next trial. Throughout both training and test, subjects were monitored to ensure that they followed instructions and did not write anything down.

Results

For all analyses reported (in both Experiments 1 and 2), the rejection level for statistical significance was set at $p = .05$.

Training performance. Figure 1 summarizes participants' mean absolute error (across the 20 training points) as a function of the presentation number (1–10) for the training points. Inspection of this figure shows that for spaced training, participants' response

[1] Due to a programming error, one item in the random massed condition was trained for eight instead of 10 trials. As there were a large number of trials, this small reduction in training for that item was unlikely to affect the results.

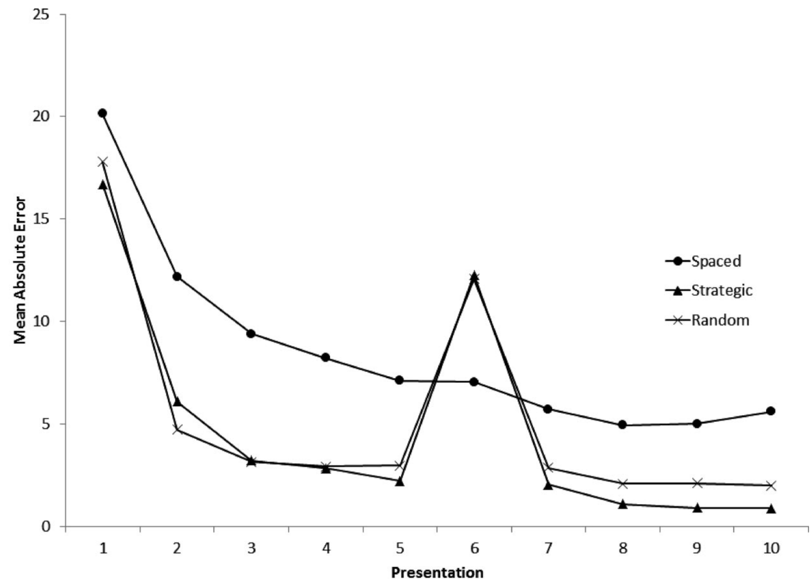

Figure 1. Mean absolute error for each presentation of training points in Experiment 1. *Strategic* = massed strategic condition; *Random* = massed random condition.

accuracy improved throughout the training period in a smooth, gradual fashion. In contrast, for both massed training groups, performance improved substantially from Presentation 1 to Presentation 2. At Presentation 6 (the second block of massed trials, which began the massed sequence over), performance dramatically worsened such that accuracy levels were nearly as low as on the very first presentation of the training points. Then, for the next presentation (Presentation 7), a massed presentation, performance quickly rebounded to near perfect levels. On the last (10th) presentation of each of the training points, responding remained more accurate after massed training (random: $M = 1.94$; strategic: $M = 0.90$) than with spaced training ($M = 5.60$), $F(2, 59) = 8.895$, $MSE = 13.75$. Still, even by the end of spaced training, the average of participants' responses was well aligned with the training function, as displayed in Figure 2.

Test performance. The mean absolute error of each participant's predictions from the actual output amounts (i.e., those generated by the function) was computed for the three types of test points: trained, interpolation, and extrapolation (see Figure 3, upper, middle, and lower panels, for mean predictions for the

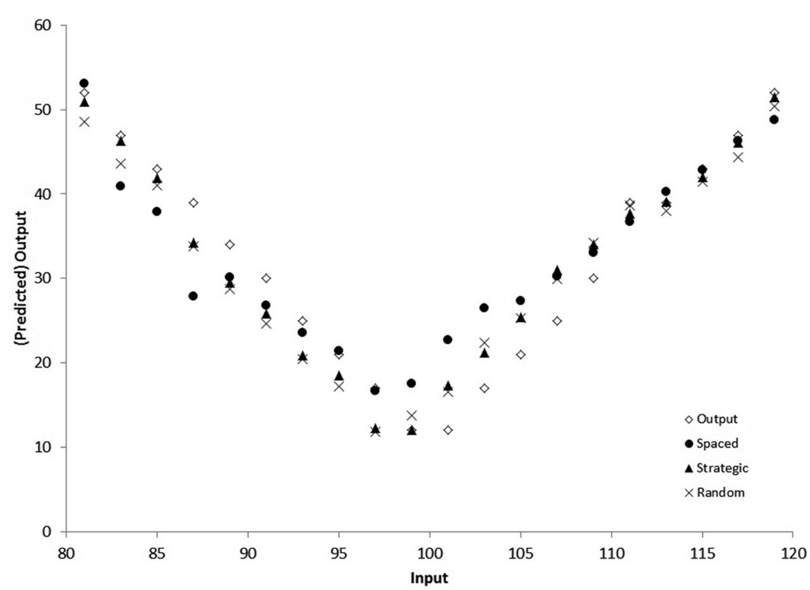

Figure 2. Mean responses for each group for the last presentation of each training point during the training phase in Experiment 1. *Strategic* = massed strategic condition; *Random* = massed random condition.

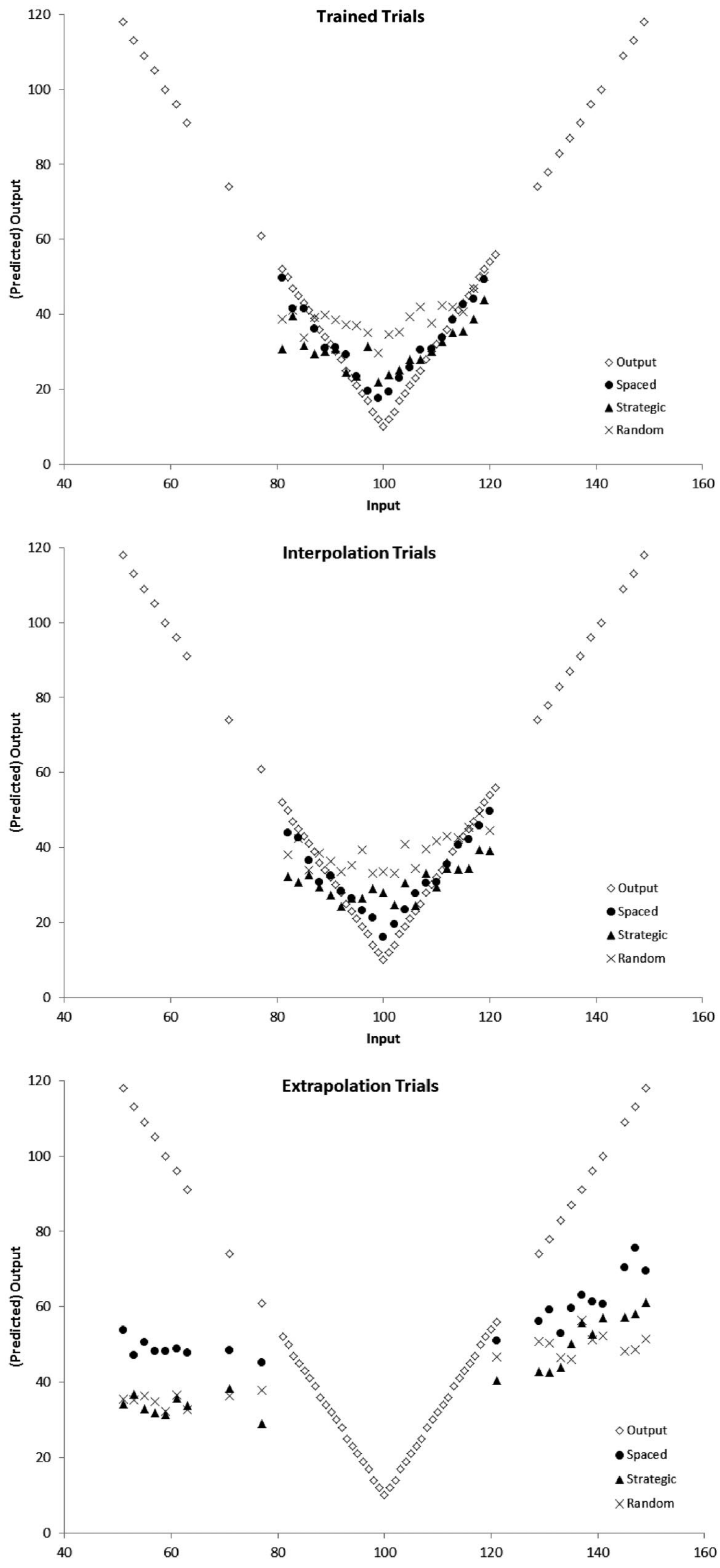

Figure 3. Mean responses for the test phase for each type of test trial (trained, interpolation, extrapolation) in Experiment 1. *Strategic* = massed strategic condition; *Random* = massed random condition.

trained, interpolation, and extrapolation points, respectively). We submitted these values to a 3 (training condition) × 3 (type of test point) mixed analysis of variance (ANOVA), with training condition as the between-subjects variable and test point as the within-subjects variable. In general, participants' responses were significantly more accurate for the trained points (M = 10.66) and interpolation points (M = 11.43) than for extrapolation points (M = 49.53), $F(2, 114) = 253.00$, $MSE = 117.11$ (for the main effect). More importantly, there was a significant effect of training condition, $F(2, 57) = 10.55$, $MSE = 182.50$, with spaced training generally producing more accurate performance (M = 17.82) than the massed strategic condition (M = 24.77) and the massed random condition (M = 29.04).

These group differences did not interact with type of test point ($F < 1$). However, because of the theoretical import of identifying the outcomes of spaced versus massed training for retention of training points and for learning of the functional relation (function form), we conducted a series of planned pairwise comparisons contrasting performance in the spaced condition relative to each of the massed conditions for each kind of test point (trained, interpolation, extrapolation). The planned comparisons confirmed that the spaced-training condition was significantly ($p < .05$) superior to the massed-random condition for the trained points, $t(38) = 2.54$; interpolation points, $t(38) = 2.73$; and extrapolation points, $t(38) = 2.82$. The spaced-training advantage over the massed-strategic condition was significant for trained points, $t(38) = 3.66$, and interpolation points, $t(38) = 4.41$, with a marginally significant difference on extrapolation points, $t(38) = 1.79$, $p = .08$.[2]

Another gauge of participants' learning of the function is the topography of their overall set of predicted values (see Figure 3). To capture the topography, we computed the slopes of each participant's predicted values for each of the three types of test points (trained, interpolation, extrapolation) for each segment of the function (left—negatively sloped; right—positively sloped). To allow straightforward interpretation of the statistical analyses, we retained the signs of the slopes and computed a 3 (training condition) × 3 (type of test point) mixed ANOVA for the left side of the function and a parallel ANOVA for the right side of the function. The ANOVAs revealed that the response slopes significantly differed across the test-point types, $F(2, 114) = 23.80$, $MSE = 0.49$, and $F(2, 114) = 15.75$, $MSE = 0.37$, for left and right segments, respectively. The mean slopes are displayed in Table 1 and clearly show that slopes were steeper for trained and interpolation points than for extrapolation points. More importantly, the slopes of participants' predictions were in general steepest after spaced training, $F(2, 57) = 7.21$, $MSE = 2.12$, and $F(2, 57) = 7.67$, $MSE = 1.48$, for the left and right parts of the function, respectively. Finally, training condition and type of test point significantly interacted such that for both the left and right parts of the function, the advantage of spaced training was especially prominent for trained and interpolation points, $F(4, 114) = 4.09$, $MSE = 0.49$, and $F(4, 114) = 2.88$, $MSE = 0.37$ (for the left and right sides, respectively).

Pairwise comparisons supported this interpretation of the interaction. The spaced condition showed a significantly ($p < .05$) steeper slope than the massed random and the massed strategic conditions on trained points—for massed random, $t(38) = 4.84$, left side, and $t(38) = 2.91$, right side; for massed strategic, $t(38) = 4.11$, left side, and $t(38) = 2.04$, right side—and on interpolation points—for massed random, $t(38) = 3.93$, left side, and $t(38) = 3.60$, right side; for massed strategic, $t(38) = 3.20$, left side, and $t(38) = 4.28$, right side. By contrast, few significant differences emerged for extrapolation. For the left (negatively sloped) side, though spaced ($M = -0.19$) but not massed random or massed strategic (random: $M = 0.08$; strategic: $M = 0.02$) training supported negatively sloped predictions, these values did not reliably differ (for extrapolation; ts < 1). For the right (positively sloped) side of the function, all groups supported positively sloped prediction. Nevertheless, pairwise comparisons indicated that spacing produced significantly more positive slopes ($M = 0.81$) in extrapolation than did the random massing ($M = 0.10$), $t(38) = 2.38$, but only nominally more positive slopes for strategic massing ($M = 0.64$; $t < 1$).

Next, to determine whether participants' responses reflected any slope at all (i.e., indicated some appreciation of the general function form), we contrasted the slopes in each training condition to zero (for each type of test point in the left and the right sides of the function). Table 1 provides the statistical results. These results showed that for spaced training, in every part of the tested function except the left (lower) extrapolation region, learners' predictions represented a slope that was significantly different from zero and in the appropriate direction (i.e., negatively sloped for the left region). After massed training, learners' response slopes were generally above zero for the right (upper) segment, but displayed slopes were not significantly different from zero for the entire lower segment (with the exception of trained points for the massed strategic group).

Discussion

One important result was that during training, massed presentations produced responding that was extremely accurate and significantly more so than for spaced presentations. The striking exception to this pattern was for the sixth presentation of the training points, the presentation that began the second block of massing (the previous massing was interrupted on this presentation). For this presentation, accuracy in the massed (but not the spaced) conditions approached the low accuracy levels that were displayed on the very first presentation of the training points. In addition, output responses to training points during the test phase were less accurate after massed training relative to spaced training. These patterns strongly suggest that participants in the massed-repetition groups were relying on working memory to generate their output responses during training. Working memory is considered to be involved in several aspects of concept learning (e.g., Craig & Lewandowsky, 2012); for massed repetitions, working memory could serve to accurately, but only temporarily, represent the correct input–output pairings. In the current paradigm, although we massed repetitions of a sequence of two training points, a learner would need only maintain two correct input–output

[2] One participant in the spaced training group displayed extrapolation that followed an oscillating pattern, a reasonable extrapolation from the V-shaped training points but nevertheless yielding relatively high mean absolute error (MAE) values when the participants' extrapolation outputs were compared to the intended V-function. To be conservative, this individual's data were included in the extrapolation analyses of MAE and slopes; excluding those data from the analyses yielded a significantly lower MAE in extrapolation for spaced than for the massed-strategic condition.

Table 1
Mean Slope of Responses on Test Trials as a Function of Trial Type, Training Group, and Segment in Experiments 1 and 2

		Lower segment			Upper segment		
Experiment	Trial type	Spaced	Massed: Random	Massed: Strategic	Spaced	Massed: Random	Massed: Strategic
Experiment 1	Extrapolation	−0.19 (−0.54)	0.08 (0.38)	0.02 (0.14)	0.81 (3.46)[a]	0.10 (0.51)	0.64 (4.25)[a]
	Interpolation	−1.47 (−6.75)[a]	−0.27 (−1.28)	−0.40 (−1.60)	1.62 (9.03)[a]	0.69 (3.68)[a]	0.58 (3.59)[a]
	Trained	−1.68 (−9.29)[a]	−0.34 (−1.64)	−0.55 (−2.64)[a]	1.60 (7.22)[a]	0.70 (3.30)[a]	1.02 (5.82)[a]
Experiment 2	Extrapolation	0.04 (0.19)		0.21 (1.90)	0.94 (4.72)[a]		0.77 (6.18)[a]
	Interpolation	−1.43 (−6.97)[a]		−0.09 (−0.57)	1.51 (8.39)[a]		0.44 (2.47)[a]
	Trained	−1.55 (−8.80)[a]		−0.10 (−0.60)	1.84 (14.00)[a]		0.81 (4.32)[a]

Note. Values in parentheses represent *t* values contrasting the slope against a value of 0.
[a] Slope is significantly different from 0.

pairings (determined via the feedback) to support virtually perfect performance on training trials repeated subsequent to the initial presentation (a memory load within the estimated capacity of working memory; Cowan, 1999). With regard to learning (test trial performance), however, based on research in the verbal learning domain (e.g., Craik & Watkins, 1973; Glenberg, Smith, & Green, 1977), maintenance of information in working memory per se does not necessarily foster storage of that information in long-term memory. In line with this idea, memory for the outputs of the trained points was less accurate after massed than spaced training (as revealed both by performance on the sixth training presentation when massing was temporarily disrupted and by test-trial performance).

For the reasons outlined in the introduction, the above finding is not necessarily telling with regard to whether massing promoted learning of the function rule relative to spacing. Though massing did not foster better learning of the associations between the input values and the output values than did spacing (as indexed by responding on the trained points during the test phase), massing could still help learners extract the relational structure reflected by the training points (the function rule). Accordingly, the second major set of findings was from the transfer test trials. These results indicated that the massed random training was also impaired in abstracting the underlying function and that even a strategically presented massed sequence did not foster acquisition of the underlying function to the extent that spaced training did. The provisional conclusion is that massed training did not foster comparative hypothesizing (Klayman, 1988) about the changes in outputs that corresponded to changes in inputs.

Before discussing these results in detail, we first report a second experiment that examined the possibility that the relatively poor test performances after massed training reflected an unintended advantage for the spaced training condition. In particular, in the spaced condition, the last 20 training trials exposed all 20 training points, whereas in the massed conditions, only 4 of the training points were exposed in the last 20 training trials. Thus, the delay between the last appearance of a training point and the onset of the test phase was longer for nearly all of the training points in the massed conditions relative to the spaced condition.

Experiment 2

To better equate the interval between training and testing, for the spaced condition we added a filled delay between the end of training and the beginning of testing for the spaced condition. We chose a delay that was half of the average amount of time it took the massed condition to complete the second 100 trials of training (in this experiment we used only the massed strategic condition, and for brevity we label this condition as *massed* from hereon). By doing so, approximately half of the spaced items were tested at a longer delay than the massed items, and approximately half of the massed items were tested at a longer delay than the spaced items, thereby generally evening out the test-delay across the two conditions.

In this experiment we also better equated the overlap between training and test order across the two training conditions (for the trained items). First, we adjusted the training order for the spaced condition to parallel that for the massed condition. That is, the pairs of training items and the order of the sequence of pairs in the massed condition were used to order the presentation of the training items in each block of the spaced training condition. Then, during testing each pair of points was presented intact (though the order of the pairs themselves was interspersed with transfer items). Finally, we used 50% more extrapolation trials in the test phase than in Experiment 1 (30 instead of the 20 used in Experiment 1) so that the slope estimates would be more robust (i.e., less influenced by an oddball response).

If the relative ineffectiveness of massing observed in Experiment 1 was a consequence of advantages enjoyed by the spaced condition in terms of a more favorable retention interval (shorter) or test task (random test trial presentation more similar to random training trial presentation), then massing might promote learning outcomes that exceed those observed for spacing. It is worth noting, however, that if such a pattern obtained, it likely would be most prominent for retention of the training points per se, as memory for conceptual information (the function form in this context) is typically more robust than for exemplars (the training points; Jones, Bourne, & Healy, 2012). Alternatively, if the effects in Experiment 1 are reflective of shortcomings in learning after massed relative to spaced training, then we should again observe superior outcomes for the spaced training condition.

Method

Subjects and design. Forty-two undergraduates from the subject pool at Washington University in St. Louis participated for partial fulfillment of a course requirement or $5 per half hour of participation. Subjects were randomly assigned to one of two conditions (*Spaced* [$N = 20$], *Massed* [$N = 22$]). The final test consisted of three trial types (*Trained*, *Interpolation*, *Extrapolation*) and was identical for all participants, resulting in a mixed 2 (condition: *Spaced*, *Massed*) $\times$ 3 (trial type: *Trained*, *Interpolation*, *Extrapolation*) factorial design.

Stimuli. The stimuli were similar to that of Experiment 1. Subjects were trained on a set of input–output pairs that made up a continuous V-shaped function within the interval $51 \leq x \leq 149$ with the vertex at 100. For values of $x \leq 100$, $f(x) = 230 - 2.2x$, and for values of $x \geq 100$, $f(x) = 2.2x - 210$, including only integers for both inputs and outputs. During training, all subjects were presented with 20 distinct trials (odd integers between 80 and 120) 10 times each, for a total of 200 trials during training. For individuals in the massed condition, participants received a pair of inputs 5 times each for a total of 10 trials before moving onto the next pair of trials. After all 10 pairs had been trained (100 trials), participants received the same trials again for a total of 200 trials. The pairs were chosen by randomly ordering items on each side of the vertex and giving each participant a single random order of pairs with the same order repeated in the second half of the training trials. For individuals in the spaced condition, the 20 training trials in each block were presented in the same order as the massed condition, such that each of the 10 pairs was seen once during each of the 10 blocks. All subjects within each condition received a single order of presentation.

After training, in the massed strategic condition participants received no delay and immediately began the test following training. In the spaced condition participants received a delay of 11.5 min, during which they played the game Tetris. This delay time was decided by examining the massed strategic condition in Experiment 1 and calculating the average amount of time it took for individuals to complete the second half of training and dividing this time by two. In this way, we approximately equated the average delay between training of the individual points and the test for both conditions. During the final test, subjects were presented with a total of 60 trials, 20 of these were the odd integers presented during training (*Repeated Training* trials), 10 were novel even integers within the trained interval (*Interpolation* trials), and the final 30 were outside the training range (*Extrapolation* trials; 15 odd integers between 51 and 79 and 15 odd integers between 121 and 149). Subjects were presented all 60 trials in a single random order, with the constraint that all training pairs remained intact during presentation. All input values were presented graphically through bar graphs. Subjects were presented with three bars. The leftmost bar represented the input value, the middle bar was adjustable using the arrow keys such that participants could respond and make predictions, and the final bar was used to show the correct output value (not displayed during the test phase). The remainder of the procedure was identical to that of Experiment 1.

Results and Discussion

Training performance. As in Experiment 1, we examined performance across the training trials. Figure 4 summarizes participants' mean absolute error (across the 20 training points) as a function of presentation number (for each particular input value). The pattern echoed that in Experiment 1. With spaced training, gradual improvement continued across the presentations. In contrast, with massed training, as expected performance improved substantially from Presentation 1 to Presentation 2; upon encountering the second block of massed trials (Presentation 6), performance dramatically worsened and then quickly rebounded to near perfect levels for the second massed presentation in the block (Presentation 7). For the last (10th) presentation of each of the 20 training (input) values, massed training continued to support highly accurate responding ($M = 0.48$), responding that was significantly more accurate than that produced with spaced training ($M = 4.11$), $F(1, 41) = 27.17$, $MSE = 5.09$. However, spaced-training participants' responses were still well aligned

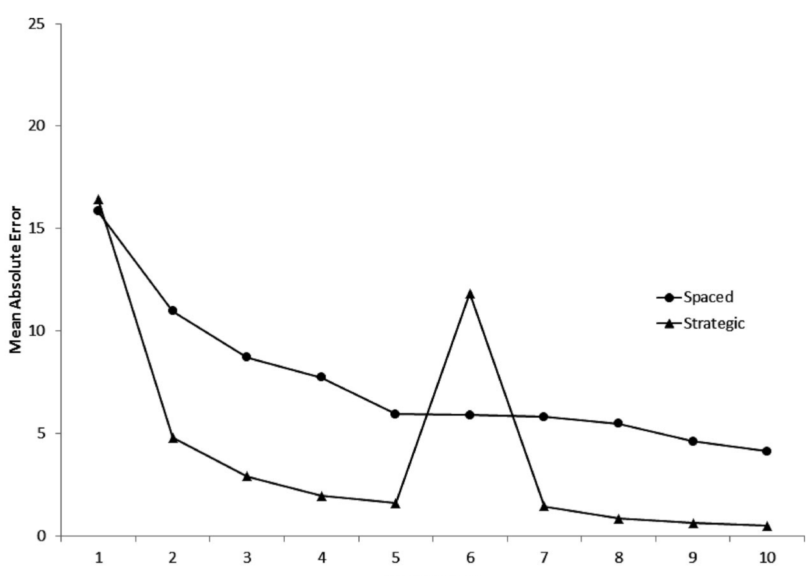

Figure 4. Mean absolute error for each presentation of training points in Experiment 2. *Strategic* = massed strategic condition.

with the training function by the end of training, as displayed in Figure 5.

Test performance. We first report the accuracy of the test responses, and then report slope analyses. The mean absolute error of each participants' predictions from the actual output amounts (i.e., those generated by the function) were computed for the three types of test points: trained, interpolation, and extrapolation (see Figure 6, upper, middle, and lower panels, for mean predictions for the trained, interpolation, and extrapolation points, respectively). We submitted these values to a 2 (training condition) × 3 (type of test point) mixed analysis of variance (ANOVA), with training condition as the between-subjects variable and test point as the within-subjects variable. In general, participants' responses were significantly more accurate for the trained points ($M = 9.44$) and interpolation points ($M = 10.64$) than for extrapolation points ($M = 41.14$), $F(2, 80) = 189.10$, $MSE = 71.53$ (for the main effect). More importantly, there was a significant effect of training condition, $F(1, 40) = 22.83$, $MSE = 123.08$, with spaced training producing more accurate performance ($M = 15.68$) in general than the massed strategic condition ($M = 25.13$). These group differences did not interact with type of test point ($F < 1$). As in Experiment 1, due to the theoretical import of identifying the outcomes of spaced versus massed training for both retention of training points and learning of the functional relation (function form), we conducted a series of planned pairwise comparisons contrasting performance in the two training conditions for each kind of test point (trained, interpolation, extrapolation).

The planned comparisons confirmed that the spaced-training advantage was significant compared to the massed condition for the trained points, $t(40) = 3.61$; interpolation points, $t(40) = 4.14$; and extrapolation points, $t(40) = 2.73$. Thus, even when (a) the delay between the most recent presentation of each of the 20 training points during training and the onset of the testing phase was on average comparable for massed and spaced training, and (b) the trained pairs presented in massed fashion were intact during test (i.e., each item of the pair was presented on successive test trials, unlike in Experiment 1), spaced training produced superior retention of the training points, as well as more accurate responses for transfer (not trained) points.

We next examined the topography of the test responses. As in Experiment 1, we computed the slopes of each participant's predicted values for each of the three types of test points (training, interpolation, extrapolation) for each segment of the function (left—negatively sloped; right—positively sloped). To allow straightforward interpretation of the statistical analyses, we retained the signs of the slopes and computed separate 2 (training condition) × 3 (type of test point) mixed ANOVAs for each segment of the function (left side, right side). The ANOVAs showed that slopes of the responses significantly differed across the test points, $F(2, 80) = 38.59$, $MSE = 0.30$, and $F(2, 80) = 5.78$, $MSE = 0.44$ (for left and right sides, respectively), such that the slopes were steeper for trained and interpolation points than for extrapolation points (see Table 1 for mean slopes). These ANOVAs also revealed that in general the slopes of participants' predictions were steepest after spaced training, $F(1, 40) = 24.83$, $MSE = 1.23$, and $F(1, 40) = 19.83$, $MSE = 0.90$, for the left and right parts of the function, respectively. Finally, a significant interaction between training condition and type of test point suggested that the advantage of spaced training was especially prominent for trained and interpolation points, $F(2, 80) = 17.24$, $MSE = 0.30$, and $F(2, 80) = 6.30$, $MSE = 0.44$ (for the interaction on the left and right sides, respectively; see Figure 6).

Pairwise comparisons supported this interpretation, with the spaced condition showing a greater slope than the massed condition on trained trials—$t(40) = 6.11$, left side; $t(40) = 4.47$, right side—and on interpolation trials—$t(40) = 5.23$, left side; $t(40) = 4.24$, right side. By contrast, no significant differences emerged for extrapolation ($ts < 1$).

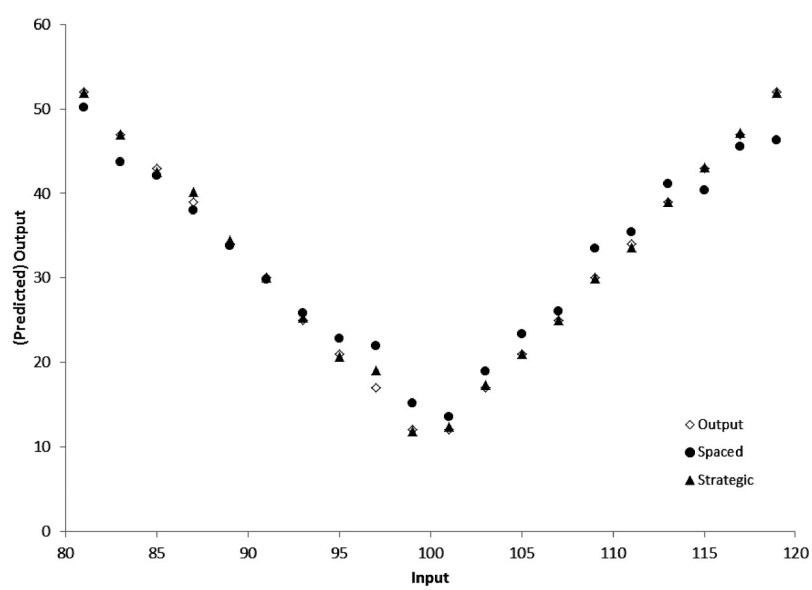

Figure 5. Mean responses for each group for the last presentation of each training point during the training phase in Experiment 2. *Strategic* = massed strategic condition.

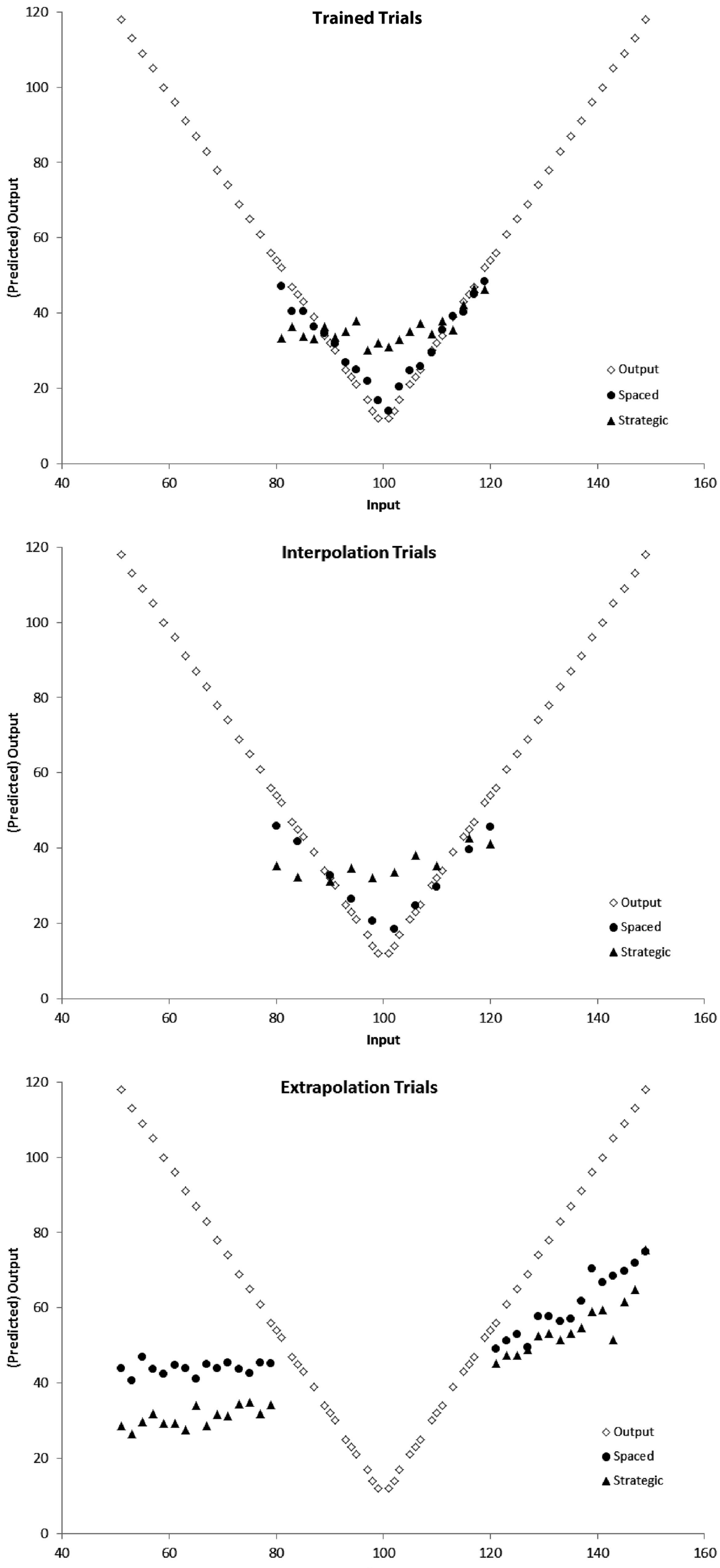

Figure 6. Mean responses for the test phase for each type of test trial (trained, interpolation, extrapolation) in Experiment 2. *Strategic* = massed strategic condition.

Next, to determine whether participants' responses reflected any slope at all (i.e., indicated some appreciation of the general function form), we contrasted the slopes in each training condition to zero (for each type of test point in the left and the right sides of the function). Table 1 provides the statistical results; these results showed that for spaced training, in every part of the tested function except the left extrapolation region, learners' predictions represented a slope that was significantly different from zero and in the appropriate direction. By contrast, after massed training, slopes were not significantly different from zero for the entire left segment; only for the upper segment did massed-training participants' predictions display a slope different from zero.

Our final analysis focused on individual participants' performances on the extrapolation points. The mean extrapolation performances displayed in Figures 3 and 6 suggest that learning about the general shape of the function was modest, at best. It is possible, though, that some proportion of participants in either the massed strategic training or spaced training did acquire reasonable knowledge about the function shape. To address this question, we examined the extrapolation mean absolute error for each individual. We conducted this analysis for both Experiments 1 and 2, and here we report the combined results. In the present function-learning task, flat extrapolation would produce a mean absolute error (MAE) of 34.72.[3] Therefore, if an individual has abstracted information about the underlying function, their MAE should be significantly less than 34.72, as they should deviate from flat extrapolation in favor of the function. Accordingly, a 95% confidence interval was calculated for each participant's extrapolation MAE, and those individuals whose confidence intervals fell entirely below 34.72 were determined to have learned the functional relationship.

After spaced training, some knowledge of the function shape was evidenced by 12 individuals (five in Experiment 1 and seven in Experiment 2), whereas with strategic massed training, only six individuals displayed knowledge of the function shape (three in each experiment; no individuals displayed knowledge of the function in the massed random training condition). This difference between the spaced groups and the strategic massed training groups ($N = 82$) was marginally significant, $\chi^2(1) = 2.95$, $p = .086$. Figure 7 provides the extrapolation profiles for these individuals (excluding the learner with sine-like extrapolation; see Footnote 3). Inspection of Figure 7 indicates that massing did not penalize those learners who did acquire knowledge of the function form, as their extrapolation profiles matched the function relatively well and were similar to that observed for the spaced-training learners.

General Discussion

Paralleling the vast literature with verbal material (Cepeda et al., 2006), the present experiments established that spaced presentations of repeated training points during a function learning task produced better learning (storage in long-term memory) of these training points than did massed presentations. This outcome was directly reflected in the test performances, with more accurate responses to the trained points after spaced than massed training. From a qualitative perspective, the superior learning achieved with spacing is also evident when examining the profile of the responses across the range of tested trained stimuli. The top panels of Figures 3 and 6 show that the responses from the spaced group captured the shape of the trained function; by contrast, the responses from the massed groups were relatively flat and indicated surprisingly little retention of the trained output values that were associated with each input value.

The transfer results in both experiments further converged on the conclusion that massed training produced significantly worse learning of the trained points, and tended to produce less abstraction of the form of the function (i.e., the intended function) than did spaced training. First, consider that participants given spaced training were able to generate output values to interpolation test points that mirrored the target function, as indicated by interpolation responses that were as accurate as responses on trained points and by slopes that were as steep as the slopes evidenced for trained points. This assessment is paralleled by visual inspection of their interpolation responses, which were virtually identical to the actual function values (see Figures 3 and 6, middle panels). Participants given massed random training, however, generated output values to interpolation points that did not reflect the target function, but were relatively invariant (across output values) and characterized a relatively flat slope (see Figure 3, middle panel); participants with massed strategic training were minimally better but still significantly and consistently (across Experiments 1 and 2) worse on accuracy and slope indices relative to participants with spaced training.

It is important to note that because interpolation can be supported by an associative learning model that acquires individual input–output associations (with some generalization around the input and the output value) but not the function rule (DeLosh et al., 1997), the interpolation patterns could, but do not necessarily, imply that spacing stimulated abstraction of the function rule per se. These interpolation patterns do reinforce that conclusion that spacing repetitions produced substantially better learning of at least the input–output associations for the cue values presented in training than did either massed condition.

The extrapolation performance more directly reveals the degree to which the three training conditions supported acquisition of the function form. Perhaps not surprisingly, random massed training (Experiment 1), for which only 30% of the trained *pairs* of input–output points represented points from one side of function's vertex, stimulated virtually no learning of the function. If participants were focusing on the particular input pairs that were massed (as intended), then it would be difficult to determine the V-shaped function from many of the training pairs presented (because they would on average reflect a fairly flat slope). Consistent with this interpretation, the extrapolation slopes were not different from zero, with fewer than 50% of the massed-training individuals displaying a positive slope on the right side of the function. It is possible that for these individuals the extrapolation profile re-

[3] In Experiment 1, one individual in the spaced training group displayed extrapolation that followed an oscillating pattern, a reasonable extrapolation from the V-shaped function. This individual's MAE (relative to a sine-like function) was compared to a criterion of 24.09 (flat extrapolation for the sine-like function), and the confidence interval for that individual indicated non-flat extrapolation. For Experiment 2, the MAE value for flat extrapolation was 35.00 (the value changed slightly with the additional extrapolation points in Experiment 2).

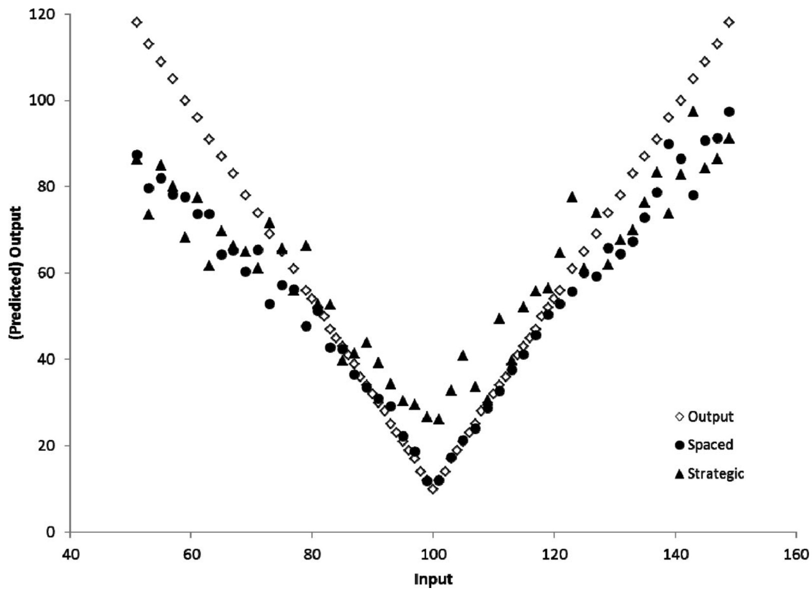

Figure 7. Mean test-phase responses on trained points (odd valued inputs from 81 to 119) and extrapolation points for participants (collapsed across Experiments 1 and 2) whose mean absolute error in extrapolation indicated some learning of the function form. *Strategic* = massed strategic condition.

flected a positive linear bias (Busemeyer et al., 1997), rather than learning of the function form.

More mixed outcomes were evident for participants with massed strategic training and with spaced training. In terms of aggregate performance, there was evidence that massed strategic and spaced-training participants learned that the right side of the function reflected a positive slope, but did not learn much about the left side of the form of the function (on average extrapolation slopes on the left side were not significantly different from zero in both Experiments 1 and 2). Converging with the aggregate performances, 90.5% of the massed strategic and 80% of the spaced-training participants (collapsed across experiments) displayed positive extrapolation slopes on the right side, whereas 21.4% and 30%, respectively, displayed negative extrapolation slopes (for the left part of the function). Even after massed strategic training, however, participants' extrapolation responses were significantly less accurate (as indexed by mean absolute error; marginally so in Experiment 1, but see footnote 2) than for participants given spaced training. These patterns clearly disfavor the theoretical expectation outlined in the introduction that massing would assist learners in learning the general function form (also, cf. Kornell & Bjork, 2008).

Importantly, this outcome obtained even though the massed procedure was modified relative to the typical massed procedure, with successive presentations of two items massed (rather than massed repetitions of one item prior to massing repetitions of the next item), and two sets of massed presentations spaced across training (rather than massing without any spacing). The massed-strategic condition was further improved by limiting the two items that were successively presented to items from the same side of the function, thereby theoretically making the slopes of the two sides of the function relatively salient. Despite our attempt to modify massing so that it would better exploit the potential for massing to convey information about the form of the function, the massed random procedure appeared to discourage learning the general function form (Experiment 1), and the massed strategic procedure did not foster better learning of the function form than did spaced training (Experiments 1 and 2).

The pattern just described obtained even when the testing sequences better matched training sequences across the massed and spaced conditions (i.e., pairs of massed stimuli were tested together) and the delay interval between training and testing was more comparable for the massed and spaced conditions (Experiment 2). Note, however, that memory for a learned function form (function concept) would not be expected to vary anyway (though memory for individual training points might) with the relatively minor differences in retention interval across conditions in Experiment 1 (cf. Jones et al., 2012). In light of this observation, when collapsing data across Experiments 1 and 2, it is telling that the percentage of participants demonstrating extrapolation performance consistent with abstracting the function rule (or at least an approximation of the function form) was double that in the spaced condition (30%) relative to the massed (strategic) condition (14%). Thus, when considering both mean extrapolation error and the proportion of participants whose extrapolation indicated some learning of the function rule, massed training fared poorly relative to spaced training in promoting learning of the function form. These patterns provide at least modest support for the conclusion that spacing of training instances is more likely to stimulate learners to acquire information about the function rule than does massed training. Most clearly, the entire set of results does not support the expectation outlined earlier that the massed training conditions would encourage most learners to focus attention on the function shape. The small proportion of massed-training learners that did evidence learning of the function shape was, however, able to learn the function form as accurately as the spaced-training learners (see Figure 7).

Theoretical Interpretations

One venerable general interpretation of the ineffectiveness of massing is that when confronted with massed repetitions, learners pay less attention to these repetitions (Hintzman, 1974; see also Cepeda et al., 2006). To explain the present deficits associated with massed training, the reduced attention would have not only undercut learning of the particular training points (consistent with the verbal learning paradigms) but also would have attenuated noticing of the relation in the changes for the outputs as a function of changes in input values for the massed pairs. However, it is not altogether clear from the attention-reduction interpretation how the strategic massing undercut noticing the relational information contained in the pairs, instead of making this relational information more salient (as intended by the modified procedure).

Another possibility is that for most learners, massing (of training pairs) per se does not promote attention to critical features needed to stimulate learners to engage in comparative hypothesizing (i.e., considering how changes in input values are associated with changes in output values). A presentation paradigm that appears to promote comparative hypothesizing is one in which the stimulus pair is presented concurrently and in which participants indicate which of the two stimuli will produce the largest (or smallest) criterion value (Pachur & Olsson, 2012). Clearly, massed training does not involve concurrent presentation; however, the nearly perfect responding during training suggests that the pairs of stimuli were concurrently represented in working memory. Accordingly, it seems more likely that massing was ineffective because, in the absence of explicit instructions to compare the stimuli, it did not itself stimulate the comparative hypothesizing necessary to learn the relational information. Indeed, massing may have disinclined participants to attempt comparison strategies, because the effort to do so would not have improved their performances during training.

This idea raises the possibility that interleaving transfer trials with training blocks, as is sometimes done in function-learning experiments (Bott & Heit, 2004; see Craig & Lewandowsky, 2012, for a similar procedure in a concept learning task), might provide an advantage for the massed condition because the task demands attention to underlying function topography (see Bott & Heit's, 2004, results). This possibility of course depends on the extent to which massing (of stimulus pairs) does provide an advantage relative to spacing for learning relational information and the extent to which massing (with interleaved transfer trials) would not encourage extensive reliance on working memory representations of input–output pairs to prepare responses for training trials. In this regard, it might be telling that even after massed participants performed poorly on the initial trial (Trial 6) of the second massed block (see Figures 1 and 4), they appeared to have not altered their approach to responding on the remainder of the training trials (as evidenced by relatively poor test performance of training trials).

The above theoretical considerations also hint that spacing of the training trials may have conferred benefits. One account championed in the verbal learning literature suggests that a spaced repetition reminds the learner of an earlier study presentation, and this reminding creates an additional trace independent from the individual study episodes. The multiple traces accruing from study and remindings especially facilitate later recall (MacLeod, Pottruff, Forrin, & Masson, 2012). For the present learning task, we would argue that spacing per se likely did not promote sufficient additional processes (e.g., remindings) to support learning. In a prior function-learning experiment, spaced repetitions in a study-only condition (participants did not have to produce output responses, but simply studied the input–output pairings) produced relatively poor learning of the training points (and poor transfer; Kang et al., 2011).

Another interpretation that seems more in line with the present function-learning paradigm rests on verbal learning paradigms in which learners must generate an appropriate solution in response to a cue word and an associated word fragment during study (Jacoby, 1978). In a massed repeated condition, the assumption is that learners rely on straightforward access from immediate (working) memory for the solution, whereas in a spaced repeated condition, learners must again "solve" the problem. The problem solving required in the spaced condition is claimed to underlie the enhanced memory for the solutions in the spaced relative to the massed condition (Jacoby, 1978). In the present function learning task, learners in the spaced condition could potentially recruit at least two sources of information to provide a "solution" of supplying an output value for a particular input value. As one source, learners might attempt retrievals from secondary (long-term) memory of previously presented points to either help construct a response for the trial at hand (DeLosh et al., 1997) or to directly recover a previous response to that particular input value. Retrieval from secondary memory would be expected to promote learning and retention of the training points (see, e.g., McDaniel, Agarwal, Huelser, McDermott, & Roediger, 2011; Roediger, Agarwal, McDaniel, & McDermott, 2011, in learning declarative knowledge; or remindings stimulated by spaced generation of responses may have stimulated additional independent encodings useful for preserving memory of the trained items, MacLeod et al., 2012). Thus, the relatively good test performance on training points (top panels in Figures 3 and 5) likely is a consequence, at least in part, of retrieving previously presented points from long-term memory to provide responses during training.

Learners could also use information about the global function form to help formulate plausible responses during training. Given that this information (how outputs changed as a function of changes in inputs) would be useful for generating a response "solution" on the training trials, learners in the spaced condition would be advantaged by attending to the relations among training points (cf. McDaniel, Waddill, & Einstein, 1988, for verbal generation effects). Along these lines, the extrapolation results suggest that some learners (30% on average) in the spaced condition did acquire information about the function form. In contrast, only 10% of the participants in the massed conditions (including random massed; 14% in the strategic massed) acquired information about the function form. We cautiously suggest that spaced training may have enhanced function-form learning, if only modestly, because that information would be useful to learners in responding during spaced training but not in massed training (because, as discussed above, for learners in massed conditions, precise input–output information for particular training points would be available in working memory). Clearly, however, spaced training (for a non-linear function) still was not optimal at stimulating the majority of participants to learn the function form.

Practical Implications

In many formal educational situations, scientific situations, and informal everyday experiences, induction of functional relations is an important learning challenge. For instance, in a textbook for a seventh grade science class attended by Mark A. McDaniel's daughter, students were provided examples of an object connected to a spring of certain tensile strength, with observations provided for different masses of the object and different distances the spring would stretch. From these observations, the students were expected to acquire the function form, as demonstrated in several extrapolation questions in which new masses were provided and the students had to generate the distance the spring would stretch. The textbook author presumably assumed that studying the examples would be sufficient for students to induce the function.

Our findings indicate that this assumption, which seems widespread (see McDaniel, 2012, for examples in other training contexts), is likely incorrect or at least limited to optimally selected presentation sequences at best. Even with college student learners, repeated study of the example training points (with spacing) does not support good learning of the function (Kang et al., 2011). Further, as shown in the present experiment, requiring learners to generate predictions from repeated massed presentation of a randomly ordered sequence of training points (the associated output for a given input) does not support good learning of the function. A new and potentially important finding, however, is that function learning is more effective with massed experiences that are strategically (rather than randomly) selected such that massed pairs are sampled from segments of the function reflecting a particular slope (Experiment 1). Still, spaced presentation of cue–criterion (input–output) points promoted better overall performance than did strategic massed training (for retention, interpolation, and extrapolation in accuracy of responses, with this pattern also generally holding for the response slopes).

Spacing may be naturally present in everyday experiences, in which experience with the cue–criterion pairs is separated across many days (e.g., relating the benefits of lawn-watering to the intensity of the sun during watering; amount of traffic congestion to the time of day) and in which people are generating predictions in order to make decisions. However, in educational settings, instructors and educational technologists need to counteract common practice and intuition favoring massed presentations and study, and perhaps consider spaced training (that requires generation of a response) to promote the learning of concepts (Kornell & Bjork, 2008) and functions. Our results also underscore that to stimulate the comparative hypothesizing presumed to facilitate function and multiple-cue learning (Klayman, 1988), educators will likely need to consider enhancing the massing versus spacing dimension highlighted in laboratory experiments with additional techniques. Possible suggestions for doing so include concurrent presentations of training stimuli, requiring comparative responses (Pachur & Olsson, 2012), and interleaving of training and extrapolation trials (Bott & Heit, 2004).

References

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

Bott, L., & Heit, E. (2004). Nonmonotonic extrapolation in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 38–50. doi:10.1037/0278-7393.30.1.38

Bourne, L. E., Jr. (1974). An inference model for conceptual rule learning. In R. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 231–256). Potomac, MD: Erlbaum.

Busemeyer, J. R., Byun, E., DeLosh, E. L., & McDaniel, M. A. (1997). Learning functional relations based on experience with input–output pairs by humans and artificial neural networks. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 405–438). Hove, England: Psychology Press.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132,* 354–380. doi:10.1037/0033-2909.132.3.354

Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). doi:10.1017/CBO9781139174909.006

Craig, S., & Lewandowsky, S. (2012). Whichever way you choose to categorize, working memory helps you learn. *The Quarterly Journal of Experimental Psychology, 65,* 439–464. doi:10.1080/17470218.2011.608854

Craik, F. I. M., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 12,* 599–607. doi:10.1016/S0022-5371(73)80039-8

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: The sine qua non for function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 968–986. doi:10.1037/0278-7393.23.4.968

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79,* 162–170. doi:10.1037/0022-0663.79.2.162

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review, 1,* 309–330. doi:10.1007/BF01320097

Foos, P. W., & Smith, K. H. (1974). Effects of spacing and spacing patterns in free recall. *Journal of Experimental Psychology, 103,* 112–116. doi:10.1037/h0036828

Glenberg, A. M., Smith, S. M., & Green, C. (1977). Type I rehearsal: Maintenance and more. *Journal of Verbal Learning and Verbal Behavior, 16,* 339–352. doi:10.1016/S0022-5371(77)80055-8

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

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17,* 649–667. doi:10.1016/S0022-5371(78)90393-6

Jones, M., Bourne, L. E., Jr., & Healy, A. F. (2012). A compact mathematical model for predicting the effectiveness of training. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Training cognition: Optimizing efficiency, durability, and generalizability* (pp. 247–266). London, England: Psychology Press.

Kang, S. H. K., McDaniel, M. A., & Pashler, H. (2011). Effects of testing on learning of functions. *Psychonomic Bulletin & Review, 18,* 998–1005. doi:10.3758/s13423-011-0113-x

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26,* 97–103. doi:10.1002/acp.1801

Klayman, J. (1988). On the how and why (not) of learning from outcomes.

In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view*. Amsterdam, The Netherlands: Elsevier.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science, 19,* 585–592. doi:10.1111/j.1467-9280.2008.02127.x

MacLeod, C. M., Pottruff, M. M., Forrin, N. D., & Masson, M. E. J. (2012). The next generation: The value of reminding. *Memory & Cognition, 40,* 693–702. doi:10.3758/s13421-012-0182-8

McDaniel, M. A. (2012). Put the SPRINT in knowledge training: Training with SPacing, Retrieval, and INTerleaving. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Training cognition: Optimizing efficiency, durability, and generalizability* (pp. 267–286). London, England: Psychology Press.

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L., III. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology, 103,* 399–414. doi:10.1037/a0021782

McDaniel, M. A., & Busemeyer, J. R. (2005). The conceptual basis of function learning and extrapolation: Comparison of rule and associative based models. *Psychonomic Bulletin & Review, 12,* 24–42. doi:10.3758/BF03196347

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (in press). Individual differences in learning and transfer: Stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General*.

McDaniel, M. A., Dimperio, E., Griego, J. A., & Busemeyer, J. R. (2009). Predicting transfer performance: A comparison of competing function learning models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 173–195. doi:10.1037/a0013982

McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language, 27,* 521–536.

Moulton, C.-A. E., Dubrowski, A., MacRae, H., Graham, B., Grober, E., & Reznick, R. (2006). Teaching surgical skills: What kind of practice makes perfect? *Annals of Surgery, 244,* 400–409.

Pachur, T., & Olsson, H. (2012). Type of learning task impacts performance and strategy selection in decision making. *Cognitive Psychology, 65,* 207–240. doi:10.1016/j.cogpsych.2012.03.003

Rea, C. P., & Modigliani, V. (1987). The spacing effect in 4- to 9-year-old children. *Memory & Cognition, 15,* 436–443. doi:10.3758/BF03197733

Reynolds, J. H., & Glaser, R. (1964). Effects of repetition and spaced review upon retention of a complex learning task. *Journal of Educational Psychology, 55,* 297–308. doi:10.1037/h0040734

Rickard, T. C., Lau, J. S., & Pashler, H. (2008). Spacing and the transition from calculation to retrieval. *Psychonomic Bulletin & Review, 15,* 656–661. doi:10.3758/PBR.15.3.656

Roediger, H. L., III, Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied, 17,* 382–395. doi:10.1037/a0026252

Rohrer, D., & Pashler, H. (2010). Recent research on human learning challenges conventional instructional strategies. *Educational Researcher, 39,* 406–412. doi:10.3102/0013189X10374770

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics practice problems improves learning. *Instructional Science, 35,* 481–498. doi:10.1007/s11251-007-9015-8

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction, 1,* 341–358. doi:10.1207/s1532690xci0103_4

Received August 27, 2012
Revision received January 6, 2013
Accepted January 14, 2013 ■