Improving Adaptive Learning Technology through the Use of Response Times

Everett Mettler (mettler@ucla.edu)
University of California, Los Angeles
Department of Psychology, 405 Hilgard Avenue, Los Angeles, CA 90095 USA

Christine M. Massey (massey@seas.upenn.edu)
University of Pennsylvania
3401 Walnut St., Suite 400A, Philadelphia, PA 19104 USA

Philip J. Kellman (kellman@cognet.ucla.edu)
University of California, Los Angeles
Department of Psychology, 405 Hilgard Avenue, Los Angeles, CA 90095 USA

Abstract

Adaptive learning techniques have typically scheduled practice using learners' accuracy and item presentation history. We describe an adaptive learning system (Adaptive Response Time Based Sequencing—ARTS) that uses both accuracy and response time (RT) as direct inputs into sequencing. Response times are used to assess learning strength and to determine mastery, making both fluency and accuracy goals of learning. ARTS optimizes spacing by expanding item recurrence intervals as an inverse function of RT. In Experiment 1, we compared ARTS to Atkinson's (1972) Markov model system using geography learning and found substantially greater learning efficiency for ARTS. In Experiment 2, we deployed the system in a real learning setting. Third graders attending an online school mastered basic multiplication facts in about two hours using ARTS, outperforming a control group using standard instruction. These results suggest that response time-based adaptive learning has remarkable potential to enhance learning in many domains.

Keywords: learning; adaptive learning; learning technology; education; instruction and teaching; memory.

Introduction

Principles of learning and memory applied to instruction might be powerfully amplified in their effects if, through adaptive learning, they can be customized to the needs of individual learners and tasks. Since pioneering work by Atkinson and colleagues (e.g., Atkinson, 1972), various adaptive learning schemes have been proposed (e.g., Pavlik & Anderson, 2008; Wozniak & Gorzalanczyk, 1994). Most systems require prior research to estimate model parameters for particular domains and learners. Sequencing is usually calculated by combining parameters, response accuracy and presentation history in a learning session.

We have developed a novel adaptive learning system (*Adaptive Response Time Based Sequencing* -- ARTS) that uses response times along with accuracy as primary inputs to govern adaptive sequencing in interactive learning. There are two primary reasons to incorporate response times in adaptive learning. First, considerable research indicates the importance of spacing in learning (for a recent review, see Pashler, Rohrer, Cepeda & Carpenter, 2007). When multiple items, categories, or procedures are to be learned, intervening intervals and/or items between presentations of a given item in a learning session can greatly improve the efficiency and durability of learning. Some important benefits of spacing relate to changing spacing as learning progresses. Using response times on interactive trials offers a more direct indicator of learning, making them a useful input into adaptive scheduling. Second, fluency itself is often a *goal* of learning. Using response times to set and meet learning criteria may offer important benefits for long term retention and fluent use of knowledge in complex problem solving situations.

Spacing and Adaptive Learning

One powerful spacing effect is that expanding intervals of retrieval practice produce better learning, relative to fixed intervals (Landauer & Bjork, 1978; Cull et al., 1996). Very recent research provides evidence for a substantial advantage of expanding the retrieval interval when material is highly susceptible to forgetting or when intervening material is processed between testing events (Storm, Bjork & Storm, 2010), conditions that apply to many formal learning situations.

Most explanations of the value of expanded retrieval intervals, and other spacing principles, involve an underlying notion of *learning strength*. Learning strength can be thought of as a hypothetical construct related to probability of successful recall on a future test. When a new item is presented, learning strength may be low, but it typically increases with additional learning trials. The value of any new test trial varies with an item's learning strength. Specifically, evidence suggests that difficulty of successful retrieval is a crucial factor (Landauer & Bjork, 1978; Karpicke & Roediger, 2007; Pyc & Rawson, 2009). Pyc & Rawson (2009) labeled this idea the "retrieval effort hypothesis": More difficult, but successful, retrievals are more beneficial. They studied the relation of number of successful retrievals to later memory performance, while manipulating the difficulty of those retrievals via number of intervening trials. Greater numbers of intervening trials led to better retention. These investigators also found evidence that, as had been suggested in other work, larger gaps produced longer average response latencies (Pyc & Rawson, 2009), a finding consistent both with the idea that a larger gap affects an item's learning strength and with the idea that learning strength is reflected in response times. One can summarize many of these findings by saying that the best time for a new presentation of an item is after the longest possible interval at which retrieval will still succeed. The idea is to stretch, but not snap, the retention interval.

Research on spacing has typically used fixed schedules, either equal intervals of item recurrence or a fixed schedule of increasing intervals. Yet different learners are likely to have different learning strengths for different items at different times, as well as differing rates of change in learning strength. Fixed schedules of recurrence cannot accommodate such variations, but adaptive learning schemes can. Previous adaptive approaches have relied on accuracy and trial history to predict learning strength, either in a Markov model estimating transition probabilities between different states of retention (e.g., Atkinson, 1972) or more elaborate models of learning (Pavlik & Anderson, 2008; Wozniak & Gorzelanczyk, 1994). Pavlik & Anderson (2008) reported strong learning results, better than with Atkinson's (1972) approach, using a detailed cognitive model of acquisition, based partially on ACT-R (Anderson & Lebiere, 1998), using prior studies to acquire learning parameters for individual items and comparable learners. Deploying such an approach in real world learning settings requires considerable up-front investment. Also, despite the value of efforts to model the learning process in exact detail, there are limits to the accuracy of any known *a priori* model. Variability among items, learners, and their interactions is substantial, requiring ongoing adjustments to the model,[1] and specific additions (such as a way to incorporate spacing effects) are needed to incorporate phenomena not originally predicted by ACT-R (Pavlik & Anderson, 2005, 2008).

Basing adaptive schemes on both accuracy and response times offers a more direct way to assess learning strength for individual learners and items in an ongoing manner. In our system, retention intervals expand as an inverse function of response time (for accurate responses), such that faster responses automatically produce longer recurrence intervals. Consistent with many studies and models, the approach assumes that learning strength is reflected in response times (Benjamin & Bjork, 1996; Karpicke & Roediger, 2007; Pyc & Rawson, 2009).

A Response Time Based Adaptive Sequencing System

Consider a set of n items (facts, patterns, concepts, procedures) to be learned. How can we implement learning principles summarized above to optimize learning of the set for the individual learner? We do so by applying principles of learning to all learning items simultaneously in a *priority score* system, in which all items are assigned scores indicating the relative importance of that item appearing on the next learning trial. Priority scores for each item are updated after every trial, as a function of learner accuracy and RTs,[2] trials elapsed, and in view of predetermined mastery criteria. Learning strength is assessed continuously and in some implementations, cumulatively, from performance data. The most straightforward version of our sequencing algorithm chooses the highest priority item for presentation on each learning trial. Adjustable parameters allow flexible and concurrent implementation of several principles of learning and memory. One important principle is that the retention interval automatically increases for an item as its learning strength grows.

In this report, we focus on item sequencing, although the system can also be applied to procedural learning and to perceptual or category learning, in which each presentation of a category involves a novel instance (Kellman, Massey & Son, 2010).

The sequencing algorithm is flexible; it may utilize any equations relating elapsed time or trials, accuracy, and RT to the priority for presentation of an item on a given learning trial. When any particular function of these variables is used, parameters may be adjusted to suit particular learning contexts and even individual learners. We describe here a characteristic priority score equation that allows implementation of several key principles of learning and has proven highly effective in our prior research. The Priority Score for item i ($\mathbf{P_i}$) is given by:

$$\mathbf{P_i = a(N_i - D)\ [b(1 - \alpha_i)\ Log\ (RT_i/r) + \alpha_i W]}$$

where:

N_i = number of trials since item i was presented
D = enforced delay constant (trials)
a, b, r = weighting constants
α_i = 0, if learning item was last answered correctly
= 1, if learning item was last answered incorrectly
W = priority increment for an error
RT_i = response time on most recent presentation of item i

Priority scores are dynamically updated after each trial. In many applications, initial priority scores are given to all items, and an item's score does not change until after it is first selected for presentation. This establishes a baseline priority for feeding in new items that may be balanced against changing priorities for items already introduced. Parameters may be set to favor recurrence of new items, items already seen, or combinations of the two.

Rapid Reappearance of Missed Items. The system ensures rapid re-presentation of items answered incorrectly by the assignment of a high priority weighting increment

[1] Because procedures for specifying these adjustments and determining numerous other parameters of the model for a new learning domain are not available in published work, we did not implement and test the Pavlik & Anderson (2008) system here.

[2] Adaptive learning systems that schedule learning events based on accuracy and speed of response are covered by US Patent #7052277. All rights reserved. For information, contact info@insightlearningtech.com.

(W). The binary variable α_i is used to activate one or the other part of the equation, depending on whether the last trial response was correct or not. If correct, α_i is set to 0, and priority becomes a function of RT. If incorrect, α_i is set to 1, and priority increment W applies to the item. With ordinary parameter settings, the error increment W will exceed all initial priority score assignments, as well as the highest priority that may result from a slow, correct answer. However, reappearance of missed items is still subject to enforced delay (see below). With typical parameter settings, a missed item will tend to have highest priority, once it passes the enforced delay.

Interleaving / Enforced Delay. To prevent presentation of an item while its answer remains in working memory (Karpicke & Roediger, 2007; Taylor & Rohrer, 2010), the system is normally configured to prevent the presentation of the same item on consecutive trials. The parameter N_i and constant D determine the enforced delay, because $(N_i - D)$ is a global multiplier in the equation. A value of 2 is typical for D, and N_i represents number of trials since last presentation of item i. Thus, the overall priority of item i will be negative on the trial immediately following the error (because $(N_i - D) = -1$). On the next trial, the priority will be 0 (because $(N_i - D) = 0$). For both negative and zero values, the priority for re-presentation of item i will be lower than all learning items having positive priority values. From then on, the priority for a missed item will be high, as its priority increment W grows proportionally to the number of elapsed trials since last presentation.

Dynamic Spacing Based on RT. The system can use various functions of RT but typically produces large priority weightings for slow, accurate responses, although not as large as for missed items. In the exemplar priority equation: For an item answered correctly, $\alpha_i = 0$, and the part of the equation involving RT is activated. RTs for inaccurately answered items are not considered meaningful. For correctly answered items, a log function of RT is used, as differences between long RTs (e.g., 20 and 30 sec) are probably not as significant as differences between short RTs (e.g., 2 and 12 sec). In this arrangement, longer spacing between presentations of an item arises automatically as the learner gives faster (accurate) responses.

Retirement Criteria. Adaptive learning focuses a learner's effort where it is needed most. We use the term *retirement* to describe removal of an item or category from the learning set, based on attainment of learning criteria. Pyc & Rawson (2007) called this "dropout" and found evidence that greater learning efficiency can be achieved with this feature, especially in highly demanding learning situations. In Exp. 1 below, the learner had to answer an item correctly and under a criterion response time on three consecutive (widely spaced) presentations to retire that item. Requiring several consecutive, fast responses to an item automatically ensures stretching of retention intervals. Thus, a retired item will have been answered quickly and accurately several times across long delays before being retired.

Our approach concurrently incorporates a number of learning principles supported by recent research. The ARTS system is built around short interactive trials, an approach supported by considerable evidence indicating that interactive "testing" trials, in which the learner makes a response, are highly effective in learning, moreso than passive presentations or "study" trials (Carpenter, Pashler, Wixted & Vul, 2008; Karpicke & Blunt, 2011). The use of systematic mastery criteria, including speed, assures both comprehensiveness and fluency in learning. As cognitive load is an important limiting factor in learning (Chandler & Sweller, 1991), it is important that items that are foundations for later learning be mastered to a reasonable degree of fluency. Finally, the rich stream of performance data accumulated by the ARTS system enables continuous assessment by instructors, and also provides several forms of learner-directed feedback, which can support specific increments in learning and sustain motivation.

Exp. 1 Comparing Adaptive Learning Systems

In Experiment 1, we compared ARTS to Atkinson's (1972) system, a classic in the literature on adaptive learning, and a benchmark against which other systems have been compared (e.g., Pavlik & Anderson, 2008). Atkinson's system was based on a Markov model tracking strength of learning items. Presentations were chosen as a function of probabilities of transitioning between three hypothetical learning states -- unlearned, temporarily learned or permanently learned. The algorithm attempted to select items that would have the highest probability of moving from an unlearned or temporarily learned state into the permanently learned state if tested and studied on the next trial. Previous learning data were analyzed to determine the model's initial parameters, including learning and forgetting rates and prior knowledge. Atkinson successfully used his model to improve learning of German-English vocabulary pairs (and used related systems in a variety of domains; for a review, see Atkinson, 1976). Performance, as measured by recall on a delayed post-test, was superior to random presentation. In the present experiment, we compared the ARTS system with a version of the Atkinson model using material that consisted of names and locations of countries on a map of Africa. To implement the Atkinson condition, item parameters were estimated using data from a previous experiment, in a manner similar to that in Atkinson (1972). No prior information was required for implementation of the ARTS system.

Method

50 undergraduates, participating for course credit, were randomly assigned to two learning conditions. One group received training using ARTS. The other group received training using the Atkinson scheduling algorithm. Each group of subjects took a pre-test in which they were asked to identify 24 countries on a map of Africa. We used countries whose location was relatively unfamiliar (e.g., Djibouti, but not Egypt). On each trial, a country was highlighted on the

map, and participants were asked to choose its name from a list of 24 country names. Countries were presented individually and no feedback was given.

The training task was identical to the pre-test, except that participants received feedback on each trial and item selection was governed by one of the two algorithms. In the ARTS condition, participants were trained until they had reached a learning criterion (responding correctly for each item three times in a row under 10 sec per item). Individual countries were removed from the learning set when retirement criteria were reached. The Atkinson system has no prescribed stopping point; we ended learning sessions after 45 minutes or a 234 trial cut-off, whichever came first. The end point was determined from pilot testing, where 234 trials was a number of trials in which more than half of participants in the ARTS condition retired all items.

Immediately after training, participants were given a post-test that was identical to the pre-test, but with countries in random order. One week later, participants returned to complete an identical delayed post-test. The entire first session took no longer than 1 hour. The experiment was run twice. The two versions were identical except that they were run on separate computers. In the first version, we discovered that the computer was introducing a delay of a few seconds between trials for the Atkinson condition. We carried out a new version with this problem eliminated. Patterns of results were indistinguishable in the two versions of the experiment, so they have been combined for this analysis.

Results

We express our primary results in terms of learning *efficiency*—post-test gains in accuracy divided by the number of learning trials invested. Adaptive response-time based sequencing produced substantially greater efficiency (53.4% greater) than the Atkinson system (Figure 1). Statistical analyses showed that efficiency was reliably higher for the ARTS condition (M=0.132) than for the Atkinson algorithm condition (M=0.086), ($t(48)= 4.33$, $p<0.001$). Post-test accuracy considered apart from learning trials invested was also reliably higher in the ARTS condition (M=0.827 vs. 0.732), ($t(48)=2.39$, $p=0.021$). A different way to view the results is to consider efficiency based on total *time* rather than trials invested (Pavlik & Anderson, 2008). Time-based efficiency (items learned per minute of training) is shown in Figure 2. In the immediate post-test, time-based efficiency for ARTS was 79% greater than in the Atkinson condition (M=0.964 for ARTS vs. 0.539 for Atkinson; $t(30)=4.50$, $p<0.001$). Values for time-based efficiency for the Atkinson condition were taken only from the subset of participants who ran on computers that were not affected by a calculation delay that added space between trials.

We carried out a separate analysis of the 1-week delayed post-test, as not every subject was tested at a delay. Participants who completed the delayed posttest (41 of 50) were included. For trial-based efficiencies, an ANOVA with factors of condition and phase showed a reliable effect of condition ($F(1,37) = 17.6$, $p< 0.001$), but no interaction ($F(1,37) = 0.811$, $p=0.371$). Efficiency for ARTS was 48% greater than the Atkinson algorithm on the delayed test, and the two conditions differed reliably (M=0.092 vs. M=0.062 respectively; $t(39)=2.09$, $p=0.043$). For time-based efficiencies, reliable differences were found between ARTS and Atkinson algorithms across tests ($F(1,37)=17.6$, $p<0.001$), with no interaction ($F(1,37)=0.81$, $p=0.370$; see Figure 2). At delayed test, the ARTS algorithm showed an 89% advantage in time-based efficiency (0.662 vs. 0.35, $t(30) = 2.78$, $p=0.009$). Response times improved from pretest to posttest but the improvement did not vary by condition.

Discussion

These results suggest that adaptive sequencing based on response times and accuracy can produce substantial enhancements in learning relative to other methods. The ARTS system was 54% more efficient on immediate post-test based on trials and 76% more efficient based on time than the Atkinson (1972) approach, and these differences were equally evident on delayed post-test. The Atkinson condition tested in this study has been shown in prior work to offer substantial improvement over random schedules of presentation (Atkinson, 1972), so we might infer that the ARTS system would outperform random schedules substantially, a prediction confirmed in other work (Kellman, Zucker & Massey, 2007).

The systems tested here differed in their prior assumptions and overall complexity. The Atkinson model, as with model-based systems in general (e.g., Pavlik & Anderson, 2008) requires pre-programming of learning parameters

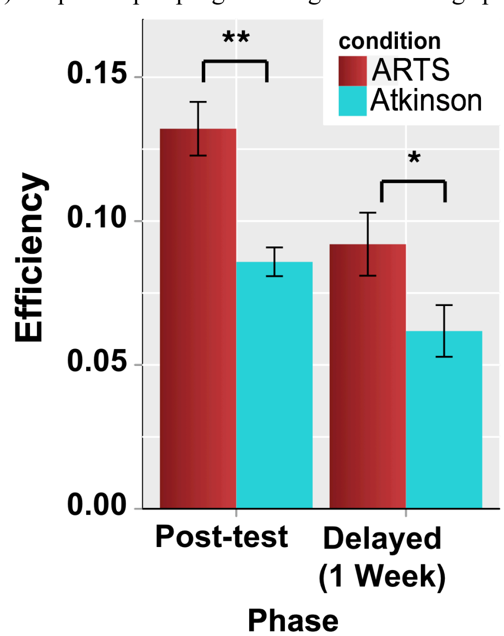

Figure 1: Efficiency for ARTS and Atkinson scheduling algorithms at immediate and delayed post-test. Efficiency equals improvement in number of post-test items answered correctly per trial of training.

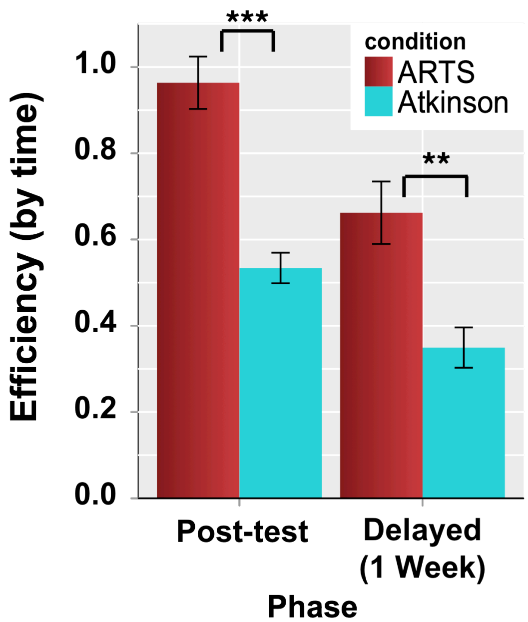

Figure 2: Time-based efficiency by test phase and scheduling algorithm. Efficiency here indicates items learned per unit time (minutes) as shown by the immediate and delayed post-tests.

based on data obtained from a prior learning experiment. With ARTS, no prior study is needed to apply the system to new domains or learners. Use of response times in interactive learning provides a more direct and up-to-date indication of learning strength as input to a sequencing algorithm.

Exp. 2 Applying ARTS to Elementary Mathematics Learning

Studying adaptive learning in genuine learning settings is crucial but has been less common than laboratory studies. One kind of challenge in real-world learning contexts is the need to do prior studies to estimate parameters in model-based systems. Another kind of challenge may be issues of diverse users, motivation, and learning materials. Students engaged in school learning may be motivated differently from paid adult subjects (as in Pavlik & Anderson, 2008), and it would be valuable to extend beyond the foreign language vocabulary used in most previous studies.

To explore these issues, we tested ARTS in a collaborative project with an online learning company that runs online charter schools in many states. We focused on third graders' learning of basic multiplication facts. Although memorization of basic math facts is one of the least appealing parts of learning in mathematics, it is a crucial foundation for later work and success in math (NCTM, 2006). Adaptive sequencing technology, we believe, can provide a highly efficient way to ensure comprehensive learning of math facts.

Method

We developed *Best Basic Math™,* an adaptive program for elementary math, and we designed a study to focus on the learning of basic multiplication facts up to 12 x 12. Specifically, 3rd grade students (n=72) in an online school in Pennsylvania logged in from home over a number of sessions in one of two conditions. Both received a pretest and posttest of 30 multiplication problems. Assessments and the learning module were web-delivered. In the treatment group (n=41), the module retained each participant's progress and current place in the learning phase across different days, and each participant's learning continued until all problems had been retired, where retirement entailed answering 4 out of the previous 5 presentations of an item correctly in less than 6.5 sec. These criteria ensured that several presentations would be widely spaced by the time any item was retired. Response time and accuracy were recorded and used in adaptive sequencing, as well as to determine item retirement. Feedback was given on each trial and also for 10-trial blocks. Overall progress toward completion was indicated at the bottom of the screen using mastery strips. For the control group (n=31), standard math lessons including multiplication content were presented as usual in the daily online sessions.

Results

For the ARTS condition, learning basic multiplication through 12 x 12 took on average 123.5 minutes (median = 109.8 min) before learning criteria were reached. Given that we were most interested in learners who had not already mastered most of this content, a primary analysis involved those students (n=28) who began with ≤ 80% accuracy on the pretest (mean pretest accuracy = 49%; mean RT = 12.6 sec per problem). Posttest scores averaged 83% accuracy and 8.3 sec per problem, gains of 69% for accuracy and 34% in fluency. Pretest to post-test gains were highly reliable for accuracy, $t(27) = 10.43$, $p < .001$, and RT, $t(27) = 5.29$, $p < .001$. Effect sizes (Cohen's d) were 2.0 (accuracy) and 1.53 (RT). The online learning company's researchers compared treatment students (n=41) with control students (n=31) who had standard assigned lessons for the same period. Groups were matched for prior performance on standardized tests. Gains of accuracy and speed for the ARTS group were highly reliable relative to the control group, $p < .01$. Effect sizes for treatment vs. control were .49 for accuracy and 1.29 for fluency. (These latter analyses did not exclude learners who were at or near ceiling on accuracy in the pretest.)

General Discussion

The studies reported here indicate that the ARTS system makes several contributions to improving the state of the art in technology-based adaptive learning systems. Specifically, in comparison to another well-known adaptive system (Atkinson, 1972), incorporating response time as a dynamic, real-time input to learning algorithms designed to implement established laws of learning and memory significantly improves the efficiency of learning. Strong learning gains were obtained in both a laboratory setting with adult learners as well as an on-line school setting with young elementary students.

The continuous stream of performance data (accuracy and speed) used in this adaptive system offers other important benefits to learning. One is the comprehensiveness of learning, based on tracking all items or categories to be learned and leading each learner to mastery criteria. In Experiment 2, about two hours of learning was sufficient to give 3rd graders reasonably complete knowledge of multiplication through 12 x 12. Although we did not study it directly here, another benefit is the use of response times in learning criteria as a means of producing fluency in learning. Finally, the rich data used by the ARTS system offers unusually rich opportunities for formative assessment and diagnosis of learning hurdles for both individuals and groups.

While the studies reported here have focused on sequencing meaningful factual *items* in mathematics and geography, the adaptive system can also be applied to other types of content, such as perceptual, category, or procedural learning. In other research, we have used adaptive algorithms to enhance pattern learning and structure extraction in high-level conceptual domains (e.g., Kellman, Massey & Son, 2010). Further, the embodiment of the adaptive system in learning technology that can be deployed without conducting prior studies to set parameters supports its potential for cost-effective application in a great variety of domains and learning settings, such as professional training in medicine, aviation, and chemistry; distance learning; and learning in K-12 schools and universities.

Acknowledgements

We gratefully acknowledge expert assistance from Tim Burke and Joel Zucker and support from IES, US Department of Education through Grant R305H06070 to UCLA. The opinions expressed are those of the authors and do not represent views of the US Department of Education.

References

Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Atkinson, R. C. (1972). Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology, 96,* 124 –129.

Atkinson, R.C., (1976). Adaptive instructional systems: Some attempts to optimize the learning process. In: *Cognition and instruction*. edited by D. Klahr, New York: Wiley. 81-108.

Benjamin, A. S. & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. M. Reder (Ed.), *Implicit memory and metacognition: The 27th Carnegie Symposium* (pp. 309-338). Hillsdale, NJ: Erlbaum.

Carpenter, S., Pashler, H., Wixted, J., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition, 36*, 438-448.

Chandler, P. & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*, 293-332.

Cull, W. L., Shaughnessy, J. J., & Zechmeister, E. B. (1996). Expanding understanding of the expanding-pattern-of-retrieval mnemonic: Toward confidence of applicability. *JEP: Applied*, *2*, 365-378.

Karpicke, J. & Blunt, J. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331(6018),* 772.

Karpicke, J. D., & Roediger, H. L., (2007). Repeated retrieval during learning is the key to long-term retention. *J. Memory and Language*, *57*, 151–162.

Kellman, P.J., Massey, C.M & Son, J. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science* (Special Issue on Perceptual Learning), *Vol. 2*, Issue 2, 285-305.

Kellman, P.J., Zucker, J. & Massey, C.M. (2007, August). *Dynamic sequencing in computer-based learning technology*. Poster presented at the annual meeting of the APA, San Francisco, CA.

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

NCTM (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence*. Reston, VA: National Council of Teachers of Mathematics, Inc.

Pashler, H., Rohrer, D., Cepeda, N., & Carpenter, S. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review, 14*, 187-193.

Pavlik, P.I. & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied, 14*, 101-117.

Pavlik, P. I., Jr., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science, 29,* 559–586.

Pyc, M. & Rawson, K. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition*, *35(8),* 1917-1927.

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

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

Taylor, K. & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, *24*, 837-848.

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