Brainscape's "Confidence-Based Repetition" Methodology

Andrew S. Cohen

Abstract

Brainscape is a synchronous web and mobile flashcard program designed to improve the retention of declarative knowledge. It is different from other spaced-repetition flashcard programs in that its pattern for re-assessment is based not on a random algorithm nor on the user's past history of correctness, but rather on the user's own judgment of confidence in each piece of information – a process that Brainscape calls Confidence-Based Repetition (CBR). In this paper, the designers of Brainscape evaluate the claim that CBR can optimize a learner's use of study time, and we highlight the large body of research that supports this claim. Our analysis concludes that Brainscape is most useful when learners have a strong intrinsic motivation to learn the topic at hand. Brainscape is particularly useful for time-starved individuals preparing for a high-stakes exam or studying a foreign language that they are very interested in learning (rather than being forced to learn).

Introduction

This paper evaluates Brainscape – a synchronous web and mobile learning application that we created to optimize the use of study time for declarative knowledge. Brainscape synthesizes the existing theories of spaced repetition and confidence-based learning to create a new technologically accessible pedagogy called Confidence-Based Repetition (CBR), which breaks declarative knowledge into its most fundamental building blocks and repeats concepts in carefully determined intervals based on the learner's confidence levels.

The need for such a convenient and pedagogically correct learning tool is epitomized by an influential, recently released U.S. Department of Education guidebook entitled "Organizing instruction and study to improve student learning". Among the guidebook's most salient recommendations are that educators (1) "Use quizzing to promote learning"; (2) "Space learning over time"; and (3) "Help students allocate study time efficiently [via metacognition]" (Pashler et al., 2007). Given the challenges of implementing these cumbersome recommendations in practice, a synchronous web and mobile tool that could automate them for both teachers and students is a welcome innovation.

The Brainscape team is especially proud of such behavioral, memorization-based innovation considering the overwhelming counter-trend toward more constructivist activities that involve a "deeper" analysis of complex systems (Uttal, 2000). Indeed, in the face of rampant criticisms that behavioral drills are deficient exercises that employ only low-level thinking and prepare learners for little more than regurgitation onto uniform examinations (Decoo, 1994), Brainscape sees itself as an important champion of behaviorism's most important tenets—presenting instruction in small steps, requiring active responses to frequent questions, providing immediate feedback, and allowing for learner self-pacing (Skinner, 1958). Brainscape helps remind us of the many cases in which behavioral study is beneficial, including cases in which the learning of rote facts *is* the educational goal (e.g. national capitals, anatomy diagrams, certain standardized test prep, or language vocabulary), and cases in which factual information first learned in constructivist environments can be reviewed using behavioral means.[1] Decoo (1994) reminds us that educators can and should still "realize drill and practice in effective and spectacular ways within even the most sophisticated [constructivist] learning environments."

The Brainscape team has designed its particular application of CBR to make independent drill and practice more efficient and thereby leave more time for constructivist, skill-based activities in the classroom. In the first section of this paper, we will analyze the efficiency of this Brainscape user experience and its unique application of CBR as a learning exercise. We will then provide a detailed analysis of why free recall, expanding

---

[1] In example of such constructivist learning could be an activity where students collaborate to paste paper cut-outs of countries onto their correct locations on a political map. While this collaborative activity may arguably be a "better" way to initially learn the map than an independent drill would be, the hypothetically stronger initial memory trace still would not guarantee permanent memorization. In this case, employing a review tool such as Brainscape could help the learner maintain her memory of the map over time.

practice, and self-regulation of study are the most important techniques to ensure long-term retention of declarative knowledge. Then, we will explore some scenarios in which Brainscape could be used in practice by individuals, teachers, or organizations. Finally, we will identify future research needed to more completely validate the Brainscape model.

I—Overview of the Brainscape Software and Experience

The goal of Brainscape's designers was to create a simple study tool for learners whose study habits are sporadic and unpredictable. Since a typical learner might study for varying lengths of time and separate her study sessions by varying intervals, Brainscape allows content creators (students, teachers, educational publishers, or Brainscape curriculum designers) to break concepts into their most fundamental building blocks that can be systematically repeated in customized intervals of time. This allows the learner to easily "pick up where she left off" without having to manually review concepts from previous sessions.

Figure 1 shows a typical "card" in Brainscape. Notice that rather than requiring a direct user response, Brainscape simply requests that the user mentally retrieve the target sentence and then manually reveal the correct answer, in the same way that she would "flip" a traditional flashcard. Brainscape then requires users to rate their confidence in the concept by answering the question "How well did you know this?" on a 1-5 scale. This Judgment of Learning (JOL) is used to determine how long until the concept is reviewed again, where higher confidence concepts are reviewed progressively less frequently.

Figure 1. Brainscape flashcards are "flipped" manually by the user; then the user enters their JOL on a 1-5 scale.

To allow the user to track her progress toward perfect confidence in a given "deck" (or a mix of several decks), Brainscape also provides several useful data visualization tools. First, the Mastery bar shows the user a weighted average of all her confidence ratings, where a deck of all un-seen cards (0s) has a Mastery of 0%, and a deck of all perfect 5s has a Mastery of 100% (the user's ultimate goal). Second, the individual bar graphs show the relative number of cards in each confidence category 0-5. Finally, the Library screen allows the user to view the average Mastery for all decks or "packages" (collections of decks) across her entire account. This diverse metacognitive snapshot provides the user with unique guidance for what subjects or concepts she most needs to study. (See Figure 2)

Figure 2. Brainscape's "Library" screen and "Stats" screens each show a snapshot of the user's confidence, in a single subject or across various subjects.

Considering that Brainscape's "flashcard"-based study experience does not require a direct user response or provide computer-generated right/wrong feedback, we have found the software to be best suited for adult learners with a strong intrinsic motivation to learn the subject at hand (such as a second language or a high-stakes standardized test). In the future, Brainscape may develop more engaging and feedback-driven widgets for younger students whose motivation (and/or metacognitive abilities) may not be as strong. We will further discuss the pros and cons of Brainscape's current feedback-light flashcard model in the Software Design Considerations section.

First, however, we will further examine the academic research that supports the principles of active recall, expanding repetition, and self-regulation upon which Brainscape is based.

II—Analysis of Study Strategies

Recall from the Introduction three of the U.S. Department of Education's most important recommendations for optimizing the organization of study: (1) "Use quizzing to promote learning"; (2) "Space learning over time"; and (3) "Help students allocate study time efficiently [via metacognition]" (Pashler et al., 2007). This chapter evaluates the underlying pedagogic theory behind each of these key strategies and helps us build a stronger theoretical base for Brainscape's flashcard engine.

A) Studying Using Prompted Recall

> *"Quizzes or tests that require students to actively recall specific information (e.g. questions that use fill-in-the-blank or short-answer formats as opposed to multiple-choice items) directly promote learning and help students remember things for longer"*
>
> --From recommendation #5 in the U.S. Department of Education's practice guide (Pashler et al., 2007)

We can all remember a time when we forgot a new acquaintance's name barely a minute after meeting them. The likely cause of this lapse is that we neglected to quietly quiz ourselves as we repeated the name aloud. ("What is his name? His name is John.") Active, prompted memory retrieval attempts could have solidified the memory trace upon each repetition.

The need for active memory recall is supported by a large body of evidence in psychology and education. Karpicke and Roediger (2006) performed a series of experiments in which participants learned lists of words and were assessed on their memory exactly one week after learning. They found that when people attempt to recall previous items during learning sessions, rather than simply “studying” them, retention was enhanced by more than *100%*. Repeated recognition-based study was conversely found to have *no* significant benefit relative to dropping items from study altogether.

Similarly, Hogan and Kintsch (1971) show that while plain study sessions (i.e. visual review) tend to be marginally better at enhancing performance on recognition tests, they are grossly inferior to retrieval practice when the end goal is to improve performance on free-recall tests. Retrieval practice should therefore be strongly recommended whenever learners truly want to know their facts.

The proven superiority of the recall method helps explain the popularity of flashcards as a study tool for many centuries. Parents, teachers, and students seem to intuitively understand that attempting to retrieve a target upon seeing a cue is the best way to learn large series of simple facts. Mobile flashcard software programs such as Brainscape can make this process more convenient by allowing learners to log many quick study sessions – and therefore many active memory retrieval events – throughout their usual daily activities, without having to worry about keeping an organized deck of physical flashcards with them at all times.

B) Spacing study sessions over time

> *"To help students remember key facts, concepts, and knowledge, we recommend that teachers arrange for students to be exposed to key course concepts on at least two occasions—separated by a period of several weeks to several months. Research has shown that delayed re-exposure to course material often markedly increases the amount of information that students remember."*
>
> --From recommendation #1 in the U.S. Department of Education's practice guide (Pashler et al., 2007)

Like the manner in which information is recalled, the temporal distribution of recall practice is a crucial determinant of the likelihood of retention. Most evidence suggests that well-spaced study sessions are almost always superior to massed sessions. Cepeda et al. (2006) performed a review 839 assessments of distributed practice in 317 experiments, and found that a whopping 96% of the cases showed a statistically significant positive effect from spacing exposure over time.

In fact, the usage of longer inter-study intervals (ISIs) has been shown to be so effective that it is even more beneficial to long-term memory retention than other factors such as verbal versus pictorial stimuli, novel versus familiar stimuli, unimodal versus bimodal stimulus presentation, structural versus semantic cue relationships, and isolated versus context-embedded stimuli (Janiszewski et al., 2003). Long ISIs also seem to have stronger benefits for verbal information and motor skills practice than they do for intellectual skills (Moss 1996). These findings suggest that the use of appropriately-spaced flashcard practice may be more efficient for studying declarative knowledge than even the fanciest of today's multimedia learning tools.

Such a strong implication demands some exploration of exactly *how long* that study sessions should be spaced apart. Pavlik and Anderson (2005) offer insight into this question using an experiment in which participants received several repetitions of Japanese-English pair recall on two different sessions, either 1 or 7 days apart. Within each of these study sessions, items to be re-tested were separated by a different number of intervening presentations: 2, 14, or 98. (The number of intervening presentations was fixed for each participant throughout the study.) The fascinating results are shown in Figure 3. Although the massed study group (receiving only 2 intervening presentations between tested items) performed better at the end of the first session when the crammed items were fresh in their minds, the study group with the greatest number of intervening presentations (98) performed best at the beginning of the second session. This was true whether the second session was one day or seven days after the first.

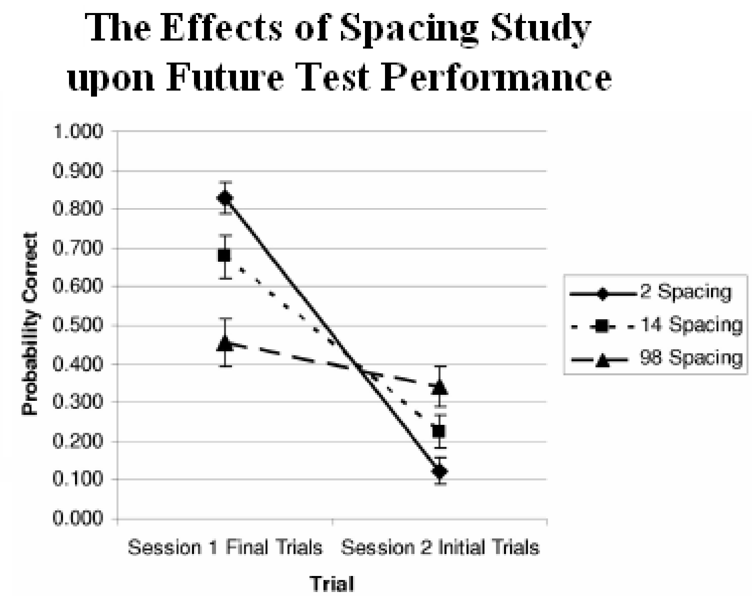

Figure 3. The longer the spacing between items in a study session, the better the performance on a recall test one week later (Pavlik & Anderson, 2005).

Nevertheless, the determination of appropriate recall intervals is not quite as simple as saying that “longer intervals lead to greater retention.” Metcalfe and Kornell (2003) show that in some cases, it may actually be advantageous to mass study together because the learning has still not yet “plateaued;” Donovan and Radosevich (1999) similarly show that intervals can sometimes be so long that the benefits from spacing begin to diminish after a certain point. Such evidence indicates that there may be some sort of middle ground between massing study and spacing study evenly.

This middle ground is known as the “expanding effect.” Proponents of the expanding effect maintain that *ISIs should be progressively increased as learners are repeatedly exposed to material*. Cull et al. (1996) performed five different experiments in which *all* showed a significant benefit for expanding practice over massed or equally distributed practice. Bahrick and Phelps (1987) and Ebbinghaus (1913) similarly propose that the best interval is the longest one before which the item is forgotten.

Figure 4 illustrates the dynamic relationship between ISI, retention interval (the amount of time before the eventual test), and memory performance across all 317 experiments included in the comprehensive literature review conducted by Cepeda et al. (2006) (see Appendix A for a summary). Not only does the graph show that spacing learning can help make retention just as good 30-2,900 days after study than it was mere seconds after study, but it shows that the longer one desires to retain a memory, the longer the optimal interval between each study session.

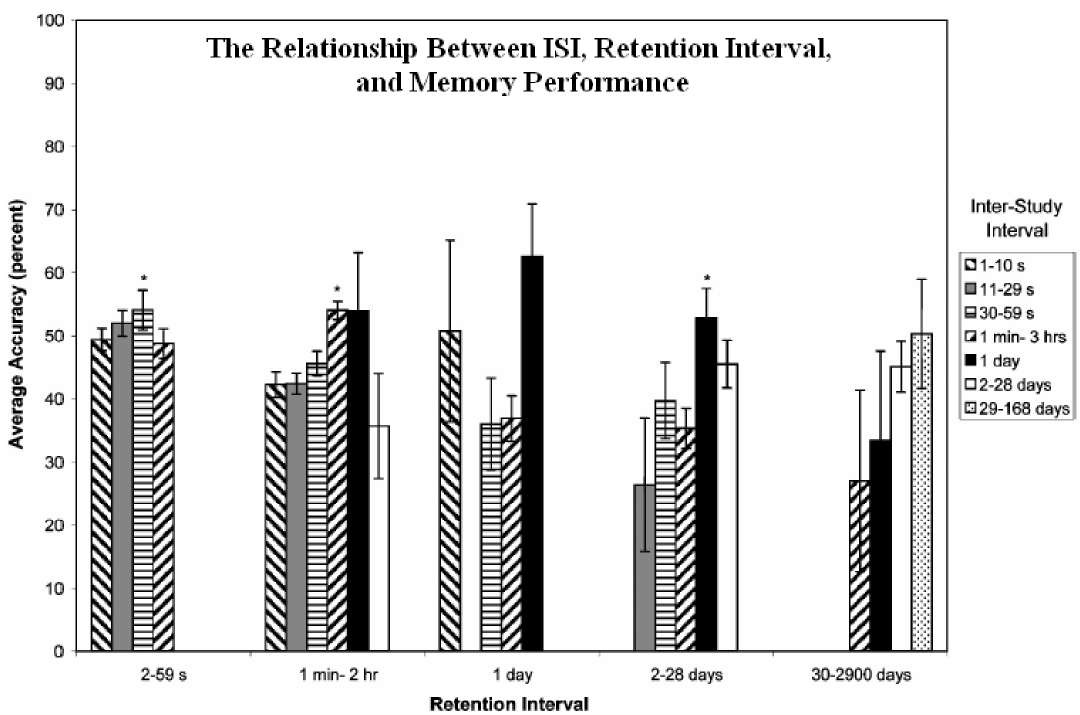

Figure 4. Note that the optimal ISI increases in step with the retention interval. If one wishes to remember something for 30-2,900 days or longer, then there is no benefit from spacing study sessions by less than 1 minute (Cepeda et al., 2006).

Spacing study sessions at increasingly longer intervals clearly appears to be the optimal method of ensuring long-term memory retention. Brainscape's designers have thus incorporated this principle into our flashcard repetition algorithm, while avoiding over-prescriptive review schedules (e.g., Super Memo) that can be discouragingly difficult to maintain for modern adults with sporadic study habits. The ordering of pending repetitions from "stalest" (i.e. least confident and/or longest interval of time since last studied) to "freshest" helps ensure that flashcard repetitions are closest to the optimal pattern without unreasonably assuming that the user has a perfectly regular study schedule.

C) Allocating Study Time Based on Metacognition

> *"To promote efficient and effective study habits, we recommend that teachers help students more accurately assess what they know and do not know, and to use this information to more efficiently allocate their study time. Teachers can help students break the 'illusion of knowing' that often impedes accurate assessment of knowledge in two ways."*
>
> --From recommendation #6 in the U.S. Department of Education's practice guide (Pashler et al., 2007)

So far, we have shown that study time is most effective when (a) items are actively recalled rather than simply reviewed, and (b) when the recall of items is performed over expanding intervals of time rather than massed at once. Yet the studies cited until this point have all used lists of items whose re-assessment intervals were determined by the experiments' *designers*. Such fixed patterns run the risk of a learner having to waste time reviewing some items that are already known perfectly, while insufficiently studying other items that need more review. Ignoring the learner's item-by-item confidence levels results in an allocation of study time that is "less than optimal" (Nelson & Dunlosky, 1991, p. 267).

For this reason a variety of researchers have set out to determine the process by which learners choose to allocate their study time. Son and Metcalfe (2000) performed a survey of 19 such studies, with 46 total combinations of treatments across different age groups, populations, experiments, or materials. They found overwhelming evidence showing that (in the absence of time constraints) *people allocate more study time to items judged to be more difficult*. Metcalfe and Finn (2007) made the same conclusions seven years later in an experiment asking participants to rate judgments of learning (JOLs) on a scale of 1-100% for several facts in a series.

The close relationship between JOL and study choice suggests that knowing how we learn best may be a natural human instinct. Indeed, Kornell and Metcalfe (2006) have performed several experiments to show that memory performance is significantly enhanced when participants are able to regulate their own study. Figure 5 shows the results of a similar experiment in which Son (2004) illustrates the interaction between JOL, study choice, and recall performance.

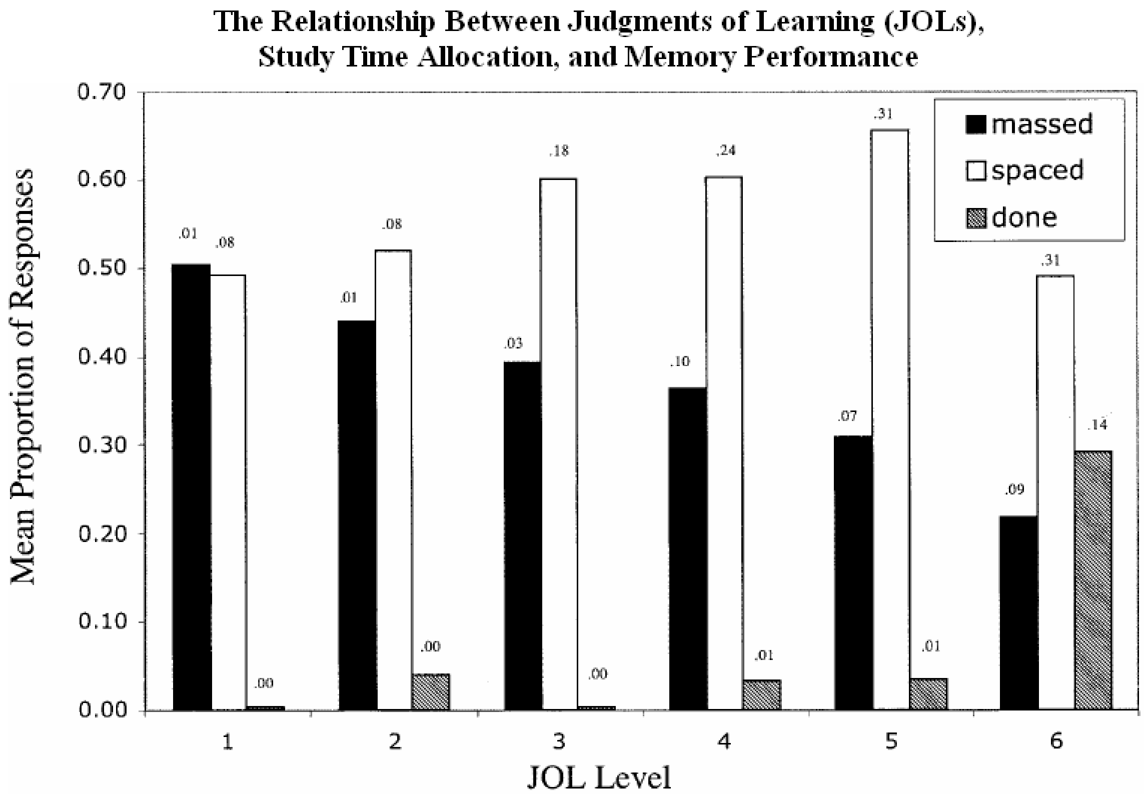

Figure 5. The better that participants judge themselves to know a particular item, the less likely they will want to study it again soon (i.e. to mass it), and the more likely they will get it correctly on a post-test (as indicated by the proportions over the bars). Participants were relatively accurate in their JOLs (Son, 2004).

In the real world, there is often insufficient time to prepare ourselves for set deadlines like exams. Such constraints suggest that we might attain short-term benefits by shifting our focus to items in our "region of proximal learning." Atkinson (1972) and Metcalfe and Kornell (2005) show that time-pressed students sometimes tend to mass their study time for these items that are "neither too easy nor too hard" in an attempt to maximize the efficiency of their sessions. This theory of proximal learning, which falls very much in line with Vygotsky's (1978) theory of "scaffolding" within the "zone of proximal development," supports the study of increasingly challenging material in increments that are just beyond a learner's current level of understanding.

Brainscape automatically enables learners to remain within their region of proximal learning by postponing repetition of “easy” items (i.e. items with a confidence rating of 5) while limiting the number of items that can exist with low JOLs before new (potentially difficult) items can be introduced into the study mix. If the number of “hard” items (i.e. confidence of 1) in the immediate rotation is approaching seven – which is the average number of items that humans are able to maintain in our short-term memory (Miller, 1956) – Brainscape will not present any new items at all until enough low-confidence items are upgraded to higher confidence. The only way to further help the learner remain within her region of proximal learning would be if Brainscape allowed her to manually “quarantine” an item that is so hard that it is not worth studying before a quickly approaching deadline. Brainscape’s designers are considering adding such a feature and allowing users to temporarily remove items from their study mix.

Whatever the presence of feedback or time constraints, the overwhelming body of research has shown that students’ performance on post-tests is improved by the ability (or encouragement) to allocate their own study time according to personal JOLs. A flashcard program that harnesses metacognition to create personalized, expanding-interval study lists would therefore be the most theoretically optimal method of preserving such declarative memory.

III—Brainscape Software Design Considerations

Throughout the design and evolution of Brainscape's flashcard engine, our designers have carefully considered the best ways to apply the aforementioned cognitive principles while preserving a web and mobile study environment that is *convenient* to the user. In this section we discuss the careful balance that Brainscape has struck between theoretical fidelity and practical efficiency.

Possibly the most fundamental early decision that needed to be made during the design process was the resolution to keep the study experience "flashcard"-like in nature. In other words, rather than requiring the user to directly input the answer to a question, to which she would receive immediate right/wrong feedback, Brainscape allows its user to simply retrieve the answer mentally and then compare her *mental* answer to the correct response that is displayed on the "back" of the flashcard. Considering the modern educational software design doctrine of requiring frequent and varied user action and providing frequent computer-generated feedback (e.g. Corbett & Anderson, 2001), this is a somewhat unconventional approach. Many educators have expressed curiosity as to whether omitting direct user feedback might risk having the learner "zone out" or to exhibit a systematic bias toward overconfident self-assessments.

Our response is simply that the possible deleterious effects of "zone out" are outweighed by the benefits of maintaining a fully learner-driven study experience.

First, we remind skeptics that the target users of Brainscape are informal, autodidactic adult learners with a distinct high-stakes learning purpose. Unlike children, highly motivated adults are naturally more likely to put effort into reflecting on their answers and managing their own progress, in the same way that diligent users of traditional flashcards are more likely than casual learners to create elaborate pile systems. For such self-directed users, Brainscape sees little need to incorporate superfluous games, animations, or other motivation enhancements simply for the sake of using such technologies.

Second, "zone out" also seems unlikely because the Brainscape software requires the user to rate her confidence level for each piece of information. This reflective not only questions the user's judgment, but also whether or not she has fully registered the piece of information. In fact, it appears that engaging in regulatory metacognitive activities, such as monitoring one's own comprehension, results in improved use of attention and other cognitive resources (Schraw & Moshman, 1995). In short, it would seem that one cannot rate his or her confidence level without paying close attention to the task at hand.

Third, the acts of self-assessment and judging one's own learning are themselves conducive to strengthening the learner's underlying memory traces. In the same way that requiring students to grade their own quizzes can help them better reflect upon their knowledge, using metacognition in a flashcard program is likely to ensure a deeper level of processing than if the program would have simply displayed whether the learner's answer was correct (Sadler, 2006). Brainscape's application of both self-assessment and progress visualization is therefore likely to deepen the learner's memory encoding while strengthening the learner's sense of mastery of the overall curriculum.

Fourth, the Brainscape team points out that the current alternatives to free mental recall are actually less effective than the basic flashcard model. Simply selecting an answer from among *multiple choices* fails to improve future performance on more meaningful active recall activities (Pashler et al., 2007; and Karpicke & Roediger, 2006), while forcing the user to *type* in an answer consumes valuable time (especially on a mobile phone) and accordingly decreases the number of repetitions that can be achieved in a given span of time. Nelson and Leonesio (1988) show that when students are separated into groups graded on either speed or accuracy, the accuracy students — despite spending significantly more time on each item — make little or no gains in performance over the speed students. This suggests that skipping the "right/wrong" step can improve study by speeding the cycle of flashcards and increasing the number of presentations in a given amount of time.

A fifth reason to avoid using the computer's judgment of a user's correctness as the determinant of future repetition intervals is that these right/wrong answers may not be an accurate representation of the user's need for repetition to begin with. For example, a careless spelling mistake could incorrectly tell the program that the user does not know an easy item, while a lucky guess could similarly mark an item as "known" even when the learner's confidence remains very low. Allowing the user to rate her own confidence rather than simply inputting an answer should therefore result in a more optimal pattern of flashcard repetition.

Sixth, encouraging self-assessment rather than computer-managed assessment may lead to significant improvements in the learner's metacognitive skills. Moreno and Saldaña (2004) and Kerly and Bull (2008) show that both children and intellectually impaired adults are able to improve their metacognitive self-assessment abilities with the help of intelligent software. Considering that normally functioning adults tend to have greater metacognitive abilities than children (Metcalfe & Finn, 2008), it is reasonable to expect that the ability to improve self-assessment skills could be even greater for adults. Black and William (1998) remind us that metacognitive reflection is among the most critical skills that any learner can develop.

Finally, Brainscape avoids directly assessing user responses in order to preserve the possibility of certain knowledge structures. For example, tasks such as visualizing the face of a given politician, providing an answer to an essay question, or humming the tone of a musical note presented on the screen in a "perfect pitch" exercise, would be much more difficult and cumbersome for a modern computer to assess. A flashcard model provides both educational publishers and individual student users with greatest flexibility in content authoring.

Given that the flashcard model was determined to be the most effective and practical for the Brainscape learning platform, the remaining questions confronting Brainscape's design team were related to *how* to best implement the confidence-based flashcard repetition experience. Let's address each of the major questions one-at-a-time.

Why does Brainscape use 5 (6) confidence categories?

Brainscape's designers chose five confidence options (numbered 1-5) because most users are already comfortable using 5-point evaluations such as Likert scales. (New, or unseen, flashcards are designated with the confidence level of 0.) This effective use of 6 categories conforms to the six normalized categories that Son (2004) employed to prove learners' preference for massing difficult items and spacing easier ones. Furthermore, we propose that providing the user with an odd number of options enables her to respond neutrally (by selecting "3"), as opposed to being forced to choose a more or less confident rating when given an even number of options.

Historically, experiments in metacognition and spaced repetition have used a broad range of methods for measuring participants' knowledge confidence. Some experimental software programs have used a simple "know/don't know" binary option, while others have provided a sliding scale from 1-100. Considering that the psychology community has considered all of these designs as academically valid, the primary remaining concern in designing Brainscape's JOL scale was simply to make it *user-friendly.*

Why does Brainscape sometimes stop showing new flashcards and only repeats existing ones?

Brainscape's flashcard repetition algorithm contains an important feature that limits cognitive load. Whenever the user has reported very low confidence in a certain number of flashcards, Brainscape stops adding new, unfamiliar flashcards into the mix until a significant number of the difficult flashcards have had their confidence upgraded. This constraint corresponds to Vygotsky's (1978) theory of scaffolding as well as Metcalfe's (2002) invocation of the theory of proximal learning, which states that learners benefit most "by directing their efforts to learning those materials that are just beyond what they have currently mastered." In other words, introducing additional difficult items before the current items are mastered could result in a cognitive overload that decreases the memory of *all* items. Brainscape refrains from introducing new flashcards unless confidence bucket 1 contains fewer than 7 items—the number that Miller (1956) states is the average that can be maintained in short-term memory.

What if people misrepresent their confidence rankings?

The concern that Brainscape users may systematically overstate their confidence is one of the largest doubts initially expressed by beginners. Yet various researchers have found that people are surprisingly accurate in assessing their memory traces. Dunlosky and Nelson (1994), for example, show that participants are able to accurately predict test performance as both an overall percentage and on an item-by-item basis, while Lovelace (1984) shows that previous exposure is not even necessary for such accuracy to prevail. According to Lovelace:

> *The memory tasks [in our experiment] involved paired-associate learning of lists of unrelated nouns and memory for sentences cued by the initial words. Probability of recall was systematically related to predictions in all conditions. Accuracy of prediction was found to increase with prior study experience with the rated material in the absence of prior test trials, although substantial prediction was possible even when predictions were made on the initial, and only, study trial. Ability to predict accurately which items would be recalled bore little or no relation to memory ability as indexed by the number of items recalled.*

Son and Metcalfe (2005) confirm these findings by showing that not only are people accurate in their self-assessments, but they are also fast. In fact, the more extreme the confidence level (i.e. "know perfectly" or "have no idea"), the faster the associated JOL is made. Brainscape users could therefore comfortably fly through most flashcards with little concern for mistaking their knowledge confidence.[2]

Despite these generalities, it is still possible that some users of Brainscape may for some reason lie about their confidence or possess inexplicably low skills of metacognitive

---

[2] Son & Metcalfe (2005) and Metcalfe & Finn (2007) show that JOLs made after a substantial delay are even more accurate because they test whether the item is in a user's long-term memory rather than simply her short-term memory. While applying such delayed JOLs to the basic version of Brainscape may be impractical, it will be considered later when we speak of the technology's future applications.

assessment. We say: Good! The eventual correction of misjudged JOLs can often yield better retention benefits than if the confidence was never misjudged in the first place. Butterfield and Metcalfe (2006) show that people are more likely to remember a corrected wrong answer when they had previously exuded high confidence that their submitted wrong answer was correct. According to this logic, if a Brainscape user fails to recall a target displaying a previously high confidence ranking, she is likely to devote more mental energies to correcting the error. Barrick and Hall (2004) show that such error corrections are even more beneficial when items are spaced rather than massed.

In fact, in a spaced or expanding environment such as Brainscape, even a systematic display of overconfidence is unlikely to hinder the user's progress. While Meeter and Nelson (2003) demonstrate that a systematic confidence bias has no effect on the *relative* proportions of items in each JOL category, Pashler et al. (2007) confirm that "the cost of overshooting the right spacing is consistently found to be much smaller than the cost of having very short spacing." Brainscape, where flashcard repetition patterns are determined based on relative rather than absolute confidence, is therefore rather immune to users' potentially poor study skills (and may help improve the users' study skills to begin with – see section II).

How can Brainscape users measure their performance?

Users can measure their performance by looking at the Mastery bar(s), which shows a weighted average of all confidence in a given deck or package (i.e. a collection of decks). Brainscape also offers a series of bar graphs representing the number of flashcards residing in each confidence bucket (0-5). Over time, the user can see her progress move from a graph where all items are in bucket 0, to a graph where all items are in bucket 5.

Kafai et al. (1998) show that the ability to visualize or quantify progress is so motivating that it frequently leads students to prefer behaviorist drilland-practice activities over the more constructionist-type activities that are favored by today's top educational theorists.

Conclusion

In this paper, we have shown that Brainscape's web/mobile learning platform conforms to the prevailing cognitive science that is necessary to ensure an efficient learning experience for declarative knowledge. It applies many of the important principles of frequent quizzing, free recall, and expanding intervals, while basing its re-assessment probabilities on the judgment of the learner herself. These features allow Brainscape's users to conveniently preserve their long-term knowledge by continuing to use the web or mobile software throughout their lifetimes.

Brainscape's applicability to such a broad range of subjects presents tremendous opportunities for the company's business development team. In its most basic form, Brainscape will develop (or import from partners) flashcard content for foreign languages, standardized tests, and other academic subjects, to be sold or distributed as stand-alone web and mobile applications. As the online content authoring environment becomes more stable and user-friendly, web users will be able to more easily add and share their own flashcards, and even export them to the Brainscape Portal mobile application, through which their web and mobile study progress will remain synchronized. The refinement of "community"-like features will then allow both learners and teachers across the globe to easily publish and share their own content for which they have particular expertise.

The expected proliferation of Brainscape users also presents exciting opportunities for data collection that will help further refine the cognitive science behind Brainscape. Researchers can use Brainscape's retail web and mobile usage data to answer questions such as:

- Which topics do users find "easiest" (based on their confidence ratings)?
- What is the average number of flashcard views before a user upgrades an item to a "5" (and how does this differ across subject areas)?
- With what patterns do users upgrade or downgrade the confidence ratings of items they have seen before?

In addition, researchers can implement Brainscape in controlled experimental settings in which learners' performance is assessed before and after the confidence-based study experience, and compared to a control group of learners who had studied for an equivalent amount of time using traditional flashcards or other study methods.

Experimental questions to explore include:

- How can the Brainscape algorithm be tweaked in order to further optimize learners' performance on posttests?
- Does creating one's own flashcards before study improve final posttest results? Is the improvement in performance significant enough to warrant the extra time spent creating one's own flashcards versus using pre-made flashcards?
- How do results of study on a computer compare to the results of studying using the identical application on a handheld device?

Whatever learning theories may ultimately be tested or improved using Brainscape, the software (as it currently stands on the market) already provides a valuable step toward making enhanced memorization techniques more accessible to today's time-starved learners. Students should consider Brainscape as a useful study tool whenever the memorization of bite-sized facts or concepts is determined to be an appropriate learning goal. Future research surrounding the Brainscape platform will serve to further explore its potential applications.

Appendix A

Summary of 317 Experiments Related to Retention Interval, Interstudy Interval (ISI), and Memory

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

Figure 6. This table shows the results of the comprehensive survey of experiments cited in the explanation of the "expanding effect" (Cepeda et al., 2006).

References

Ahlstrom, V., & Longo, K. (2001). Human factors design guide update (Report number DOT/FAA/CT-96/01): A revision to chapter 8 - computer human interface guidelines. Retrieved November 2005, from http://acb220.tc.faa.gov/technotes/dot_faa_ct-01_08.pdf.

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

Badre, A.N. (2002). *Shaping Web Usability: Interaction Design in Context*. Boston, MA: Addison Wesley Professional.

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

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

Bahrick, H., & Hall, L. (2004). The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory and Language*. Vol. 52, No. 4, 566-577.

Bailey, R.W. (1996). *Human performance engineering: Designing high quality professional user interfaces for computer products, applications and systems* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment.* London: King's College.

Butterfield, B., & Metcalfe, J (2006). The correction of errors committed with high confidence. *Metacognition and Learning,* 1(1), 69-84.

Cepeda, N., Pashler, H., Vul, E., Wixted, J., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354-380.

Cull, W., Shaughnessy, J., & Zechmeister, E. (1996). Expanding understanding of the expanding-pattern-of-retrieval mnemonic: Toward confidence in applicability. *Journal of Experimental Psychology: Applied*, 2(4), 356-378.

Czaja, S.J., & Sharit, J. (1997). The influence of age and experience on the performance of a data entry task. *Human Factors and Ergonomics Society* Annual Meeting Proceedings, 144-147.

Decoo, W. (1994). In defence of drill and practice in CALL: A reevaluation of fundamental strategies. *Computers & Education*, 23(1-2), 151-158

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

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

Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multitrial learning. *Journal of Gerontology: Psychological Science*, 52, 178-186.

Dunlosky, J., & Nelson, T.O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language*, 33, 545–565.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. New York: Teachers College, Columbia University.

Farkas, D.K., & Farkas, J.B. (2000). Guidelines for designing web navigation. *Technical Communication*, 47(3), 341-358.

Fogg, B.J. (2002). Stanford guidelines for web credibility. A research summary from the Stanford Persuasive Technology Lab. Retrieved November 2005, from http://www.webcredibility.org/guidelines/.

Galitz, W.O. (2002). *The essential guide to user interface design*. New York: John Wiley & Sons.

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

Glover, J. A. (1989). The “testing” phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, 81, 392–399.

Heinich, R. (1970). Technology and the management of instruction (Association for Educational Communications and Technology Monograph No. 4). Washington, DC*: Association for*

Educational Communications and Technology.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77–99). Hillsdale, NJ: Erlbaum.

Hogan, R., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, 10(5), 562-567.

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

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

Jonassen, D.H. (2006). *Modeling with technology: Mindtools for conceptual change*. New Jersey: Pearson Prentice Hall.

Kafai, Y.B., Franke, M., Ching, C., & Shih, J. (1998). Games as interactive learning environments fostering teachers' and students' mathematical thinking. *International Journal of Computers for Mathematical Learning*, 3(2), 149-193.

Karpicke, J., & Roediger, H. (2006). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57(2), 151-162.

Karpicke, J., & Roediger, H. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval practice enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 704-719.

Kelemen, W. L. (2000). Metamemory cues and monitoring accuracy: Judging what you know and what you will know. *Journal of Educational Psychology*, 92, 800-810.

Kerly, A., & Bull, S. (2008). Children’s interactions with inspectable and negotiated learner models. In Woolf, B., et al. (Eds.), *Lecture Notes in Computer Science* (pp. 132-141). Springer-Verlag, Berlin Heidelberg.

Koriat, A., Bjork, R., Sheffer, L., & Bar, S. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General,* 133(4), 643-656.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(3), 609-622.

Koyani, S.J., & Nall, J. (1999, November). Web site design and usability guidelines. *National Cancer Institute, Communication Technologies Branch Technical Report*. Bethesda, MD.

LaMonica, M. (2005). Ajax spurs Web rebirth for desktop apps. *CNET News*. Retrieved from http://business2-cnet.com.com/AJAX+spurs+Web+rebirth+for+desktop+apps/2100-1012_3-5977268.html.

Leavitt, M., & Shneiderman, B. (2006). Research-based web design & usability guidelines. *U.S. Department of Health and Human Services*. Retrieved from http://www.usability.gov/pdfs/ guidelines.html.

Lightner, N.J. (2003). *What users want in e-commerce design: Effects of age, education and income. Ergonomics*, 46(1-3), 153-168.

Lovelace, E. (1984). Metamemory: Monitoring future recallability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(4), 756-766.

Lumsdain, A.A., & Glaser, R. (1960). *Teaching machines and programmed learning: A source book*. Washington, DC: National Education Association.

Mager, R.F. (1962*). Preparing objectives for programmed instruction*. Belmont, CA: Fearon.

Meeter, M., & Nelson, T. (2003). Multiple study trials and judgments of learning. *Acta Psychologica*, 13(2), 123-132.

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

Metcalfe, J., & Finn, B (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review,* 15(1), 174-179.

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, 52, 463-477.

Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General*, 132(4), 530-542.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General*, 131(3), 349–363

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97

Moreno, J., & Saldaña, D. (2004). Use of a computer-assisted program to improve metacognition in persons with severe intellectual disabilities. *Research in Developmental Disabilities*, 26(4), 341-357.

Morkes, J., & Nielsen, J. (1998). Applying writing guidelines to Web pages. Retrieved November 2005, from http://www.useit.com/papers/webwriting/rewriting.html.

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

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 676–686.

Nelson, T.O., & Dunlosky, J. (1991). When people's judgments of learning (JOL) are extremely accurate at predicting subsequent recall: The delayed-JOL effect. *Psychological Science*, 2, 267-270.

Nelson, T.O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, 5, 207-213.

Nielsen, J. (2003, November 10). The ten most violated homepage design guidelines. Alertbox. Retrieved November 2005, from http://www.useit.com/alertbox/20031110.html.

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). Organizing instruction and study to improve student learning. Institute for Educational Sciences practice guide, *U.S. Department of Education*.

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

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

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

Reiser, R., & Dempsey, J. (2007). *Trends and issues in instructional design and technology*. New Jersey: Pearson Prentice Hall.

Sadler, P. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1), 1-31.

Schraw, G. & Moshman, D. (1995). Metacognitive Theories. *Educational Psychology Review*, 7, 351-371.

Shaughnessy, J. J., Zimmerman, J., & Underwood, B. J. (1972). The spacing effect in the learning of word pairs and the components of word pairs. *Memory & Cognition*, 2, 742–748.

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 907–912.

Sisti, H., Glass, A., & Shors, T. (2007). Neurogenesis and the spacing effect: Learning over time enhances memory and the survival of new neurons. *Learning and Memory*. Vol. 14, 368-375.

Skinner, B.F. (1958). Teaching machines. *Science*, 128, 969-977.

Slamecka, N., & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Pscyhology: Learning, Memory, and Cognition*, 9(3), 384-397.

Son, L. (2004). Spacing one's study: Evidence for a metacognitive control strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(3), 601-604.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 204-221.

Son, L. K., (2002, November). *Metacognitively-controlled spacing of study*. Paper presented at the annual meeting of the Psychonomic Society, Kansas City, MO.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 204–221.

Son, L. K., & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage process. *Memory and Cognition*, 33(6), 1116-1129.

Sperling, G.A. (1967). Successive approximations to a model for short-term memory. *Acta Psychologica*, *27*, 285-292.

SuperMemo (Not a Mission Statement) (n.d.) Retrieved Feb 12, 2008 from Japan Today web site: http://www.japantoday.com/forum/printable.asp?m=697576&mpage=2

SuperMemo FAQ (n.d.) Retrieved Feb 12, 2008 from SuperMemo web site: http://www.supermemo.com/help/faq/begin.htm

Uttal, W. (2000). *The war between behaviorism and mentalism: On the accessibility of mental processes.* London: Lawrence Erlbaum Associates.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*, 79-91. Cambridge, MA: Harvard University Press.

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

Wilson, J.R. (2000). The place and value of mental models. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 1-52.

Woodworth, R. S. (1938). *Experimental psychology*. Oxford, England: Holt.

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