The Cognitive Science of Learning Enhancement:
Optimizing Long-Term Retention

Professor Eric Eich, Department of Psychology ee@psych.ubc.ca   BIO

Everywhere you look around UBC, there are signs that the university is serious about raising its already high standards for faculty teaching and student learning. One indicator is the Carl Wieman Science Education Initiative--an ambitious project, led by physics professor and Nobel Laureate Carl Wieman, that seeks to systematically improve and scientifically measure the effectiveness of undergraduate science education. A second example is the Centre for Teaching and Academic Growth: from its modest beginnings in 1987 (see Riddell, 2007), TAG has evolved into a major program that helps faculty members and graduate students hone their teaching skills through a host of workshops, support services, and professional development activities. Other developments of note include the LEAD initiative (designed in part to assist UBC's educational units in planning and fulfilling their own measures of educational success), the Postdoctoral Teaching Fellow program (funded by the President's Office to promote teaching in the faculty of Arts), and the 2008 debut of UBC's Celebrate Learning Week (a welcome complement to the university's more traditional Celebrate Research Week).

In addition to these broad initiatives, significant innovations in teaching and learning are taking place in many individual departments, schools, and programs on both UBC campuses. Thus, for example, students in Peter Newbury's introductory course on astronomy build a human orrery--a machine in which the planets in our solar system orbit the sun when you turn a crank. In Newbury's case, students play the part of the planets and walk around a model Sun at varying speeds--a novel and engaging way for them to appreciate celestial mechanics. These same adjectives apply to the Carol-Ann Courneya's "Heartfelt Images" photography project, in which first-year medical/dental students create images (still or video) that capture the aesthetic or conceptual essence of what they are learning about the heart and blood vessels.

Academic units as varied as AHVAT (Art History, Visual Art, and Theory), Cellular and Physiological Sciences, Dentistry, and Educational Studies are placing a premium on self-learning and interactive methods of teaching. At the same time, a similarly diverse set of units, including Marking and Political Science, are searching for practical, meaningful ways to get their students out of the classroom and into the real world, or to bring the community into the classroom on campus, all with a view to providing a richer, better-rounded educational experience. Several departments, including History and Psychology, have revamped their undergraduate curricula to provide more hands-on research activities for their students, while several other units are pursing technological approaches to improve teaching and learning. For instance, the Faculty of Land and Food Systems has created a web-based simulation, called "Green Genes," to aid students' understanding of genetics. Meanwhile, researchers in Electrical and Computer Engineering have developed state-of-the-art wireless devices to allow instructors complete mobility as they lecture, so that they can write, talk, and project from anywhere in the classroom.

Most of these local, unit-based innovations were described in the Summer 2009 issue of Tapestry: a special edition, which more than 20 UBC faculty members from both campuses shared their ideas and insights about teaching. In this issue of Tapestry, I wish to continue the conversation, but take it in a new direction.

Recent years have witnessed an explosion of interest in bringing advances in cognitive science out of the laboratory and into a variety of educational settings (elementary school classrooms, university lecture halls, Web-based tutorial systems, etc.). The spark for much of this interest has been the Cognition and Student Learning program (CASL), an initiative of the Institute of Education Sciences within the US Department of Education. Established in 2002, the program aims to produce "an array of tools and strategies (e.g., instructional approaches, computer tutors) that are based on principles of learning and information processing gained from cognitive science and that have been documented to be efficacious for improving learning in education delivery settings from prekindergarten through high school and for vocational or adult basic education or developmental (remedial) … programs for under-prepared college students" (http://ies.ed.gov/funding/ncer_rfas/casl.asp). To that end, CASL sponsors translational research on such wide-ranging topics as:

- applying cognitive principles of reasoning to build practical lesson plans for elementary school students (Klahr & Li, 2005),
- advancing the math skills of low-achieving adolescents in technology-rich learning environments (Bottge, Rueda, & Skivington, 2006),
- dynamically modifying the learning trajectories of novice chemistry students (Beal, Qu, & Lee, 2008),
- improving young children's numerical understanding (Richland, Holyoak, & Stigler, 2004; Siegler & Ramani, 2006),
- identifying the neural markers of effective learning (Anderson, 2007),
- promoting abstract thinking in kindergartners to jump-start their acquisition of academic content knowledge (e.g., reading and math; Pasnak, Cooke, & Hendricks, 2006),
- recognizing and rectifying 6th graders' scientific misconceptions of science (Heckler, Kaminski, & Sloutsky, 2008), and
- sharpening children's metacognitive skills and control strategies, so that they can more effectively allocate and organize study time and effort optimally (Metcalfe & Kornell, 2003).

In the remainder of this essay, I focus on yet another issue that has captured the attention of many CASL-sponsored investigators: how to optimize long-term retention. Interest in this issue has been triggered by two things. One is the idea, proposed by Pashler (2006), that educational failures often reflect problems in retaining information over time, rather than in acquiring the information in the first place. Placing the onus on forgetting makes it easier to understand why one-third of American youths can't locate the Pacific Ocean on a map, why the majority of Harvard undergraduates can't explain why it is colder in winter than summer, or other egregious gaffs. As Pashler (2006, p.26) has pointed out, "Progress in finding out what mitigates forgetting should be helpful not only in school, but in job training as well." At the same time, a better understanding of what mitigates forgetting would complement the rich literature on the acquisition and comprehension of knowledge--issues that have occupied educational psychologists and teaching professionals for decades (e.g., Anderson & Faust, 1973; Bransford & Donovan, 2005; Lesgold & Glaser, 1989).

The second trigger is the counterintuitive finding that interventions that seem to make learning more difficult and slow the rate of learning can actually be effective in enhancing long-term retention. The benefits of these desirable difficulties (Bjork, 1994) were first observed in laboratory experiments that achieved high levels of precision and control, but at the cost of low ecological validity. However, more recent studies suggest that the same benefits accrue under more realistic, educationally meaningful conditions. In the following sections we will review

several of these studies, along with other examples of contemporary cognitive research that speaks to both the science and practice of learning enhancement.

Learning, Remembering, and the Spacing of Practice

One of the most a venerable, robust, and reliable phenomenon of human memory is the spacing effect: the observation that increasing the temporal interval between successive study episodes often enhances performance on a later memory test. The phenomenon was established by Ebbinghaus (1885/1964) in his seminal studies of savings in the relearning of nonsense syllables--the first experimental investigations of human memory. Since then, hundreds of additional spacing-effect studies have been published (Pashler, Rohrer, Cepeda, & Carpenter, 2007), many of which followed the basic design depicted in Figure 1. By this design, participants have two opportunities to learn the same set of to-be-remembered or target items (words, pictures, stories, etc.). These study episodes are separated by a varying gap or interstudy interval (ISI), while the gap between the second study episode and the final test on the material represents the retention interval (RI). Most experiments incorporate several values of ISI and a single, fixed value of RI.
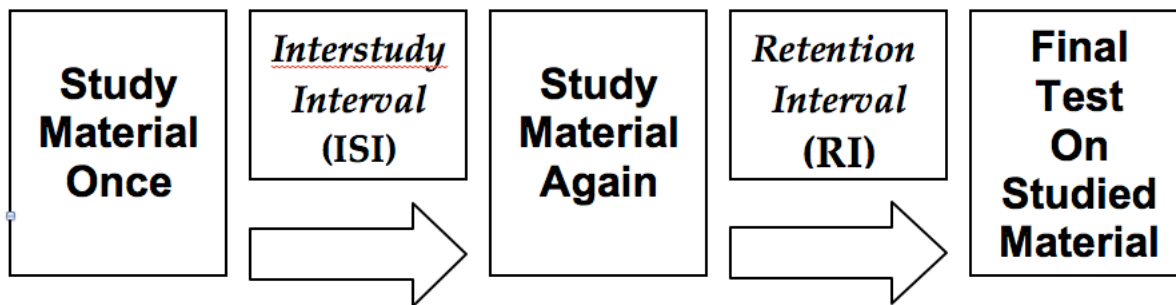

Figure 1: Typical design of an experiment on the spacing effect. Participants study to-be-remembered or target material on two occasions, separated by an interstudy interval (ISI). Their memory for the material is tested for memory after a retention interval (RI) measured from the end of the second study occasion. Most spacing-effect experiments have one RI and several values of ISI. Source: H. Pashler et al. (2007), Psychonomic Bulletin and Review, 14, 187-193. Used with permission.

Consistent with Ebbinghaus, the nearly ubiquitous finding has been that distributing study episodes over time leads to better long-term retention than does continuous or massed practice, even when the ISI is as short as a few minute (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). The positive effects of spacing are large and they are evident in many domains, including classical conditioning, verbal learning, picture memorization, text comprehension, and motor skill acquisition (Goettl, 1996). What is more, the spacing effect is not unique to humans: C. elegans, a tiny but talented nematode with only 302 neurons, shows enhanced long-term retention of a habituation response following distributed as opposed to massed training trials (Beck & Rankin, 1997; Rose, Kaun, & Rankin, 2002). So pervasive is the phenomenon that Ulrich Neisser--the Cornell professor who coined the term "cognitive psychology" in his eponymous 1967 book--was moved to write a poem about it. After listening to a conference presentation by the subject by Robert Bjork (1988, p.399), Neisser wrote:

You can get a good deal from rehearsal,
If it just has the proper dispersal.
You would just be an ass

to do it en masse:
Your remembering would turn out
much worsal.

Though the rich data base on spacing has animated much discussion among cognitive theorists (Baddeley, 1990), it has drawn attracted little interest among education experts. This was true twenty years ago--witness Frank Dempster's (1988) article in the American Psychologist titled "The spacing effect: A case study in the failure to apply the results of psychological research"-- and it remains true today. As Pashler and his associates have recently remarked:

> Whether one looks in classrooms, instructional design texts, or at current instructional software, one finds little evidence that anyone is paying attention to the temporal distribution of study. Moreover, programs that deliberately compress learning into short time spans (immersion learning, summer boot camps) seem to be flourishing (Pashler et al., 2007, p.188).

Why do educators continue to give short shrift to the spacing effect? A recent review paper by Cepeda et al. (2006) provides a compelling answer. They note that although the cognitive literature on spacing effects is vast, with over 400 empirical publications, only about 3% of these reports involved retention intervals as long as 1 day, while a scant 1% delayed the final memory test more than 1 week. Given these figures, Cepeda and his colleagues argued that:

> Although psychologists have decried the lack of practical application of the spacing effect …, the fault appears to lie at least partly in the research literature itself: On the basis of short-term studies, one cannot answer with confidence even basic questions about the timing of learning. For example, how much time between study sessions is appropriate to promote learning and retention over substantial time intervals? Is it a matter of days, weeks, or months? (Cepeda et al., 2006, p.1095).

Recent research by a team of researchers, led by UCSD's Hal Pashler, is shedding light on this question. To increase the generality of their results to practical contexts, Pashler and his co-workers use materials that are broadly representative of the sorts of cognitive challenges people meet in everyday life--for instance, learning foreign-language vocabulary, absorbing new factual knowledge, or acquiring a new mathematical skill.

In one set of experiments (see Cepeda et al., 2006; Pashler et al., 2007), participants first learned 40 paired associates, each consisting of a Swahili word (somo, for example) and its English equivalent (friend). Participants continued to study the list until they could correctly translate each foreign word on two separate occasions, to ensure their mastery of the materials. Immediate, corrective feedback was given each and every time an error was made. A second session was held either immediately after the first or following a variable delay. During this re-study session, a fixed number of additional learning trials were given on the same word pairs. A final test of retention (somo - ????) occurred 10 days after the second study session. Consistent with prior reports, the results revealed a marked improvement in final-test recall as the interstudy interval increased from 15 minutes to 1 day. However, final-test performance decreased by a small amount as ISI increased beyond 1 day.

In another series of studies (Cepeda et al., 2006; Pashler et al., 2007), participants were taught little-known facts (e.g., Rudyard Kipling invented snow golf), as well as the names of seldom-seen visual objects (Amelia Earhart made her ill-fated flight in this model of Lockheed

Electra). Participants studied these same bits of arcana some later time, and were given a final test of recall six months after the second study session (Who invented snow golf? Name this model airplane, in which Amelia Earhart made her ill-fated last flight.). Under these conditions, final-test performance improved until the interstudy interval reached about one month, with a small and shallow drop off after that. Similar results obtained in a third series of studies (see Rohrer & Pashler, 2007; Rohrer & Taylor, 2006) in which participants learned to solve problems in combinatorics (permutations and combinations)--a more abstract form of learning than that required to memorize concrete facts.

Taken together, the research of Pashler and his colleagues demonstrate that "powerful spacing effects occur under practically meaningful time periods" (Rohrer & Pashler, 2007, p.185) and that these effects emerge within practically meaningful contexts involving the acquisition of new facts, foreign languages, or mathematical skills.

That said, how should one distribute his or her study time to optimize long-term retention? Interestingly, the answer is: it depends on how long you wish to remember something (Cepeda et al. 2006). In reference to Figure 1, the data indicate that as the retention interval increases, the optimal interstudy interval also increases, while the ratio of optimal ISI to RI decreases (Cepeda et al., 2006; Cepeda, Coburn, Rohrer, Wixted, Mozer, & Pashler, 2009; Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008). Generally speaking, the window for optimal memory performance appears to open when the interstudy interval is between 10 to 20 percent of the retention interval. Thus, a one-day gap between initial study and review is suitable for a one-week retention interval, while students should restudy material learned at the beginning of the fall semester around the end of September, in order to maximize their retention of a final exam in December. For even longer retention intervals, Cepeda et al. (2006, p.1101) advise that:

> If a person wishes to retain information for several years, a delayed review of at least several months seems likely to produce a highly favorable return on investment--potentially doubling the amount ultimately remembered compared with a less temporally distributed study schedule, with study time equated. Although this advice is in agreement with the earlier work of Bahrick (e.g., Bahrick et al., 1993 [one of the few early cognitive experiments that used practically meaningful study and test intervals], it is at odds with many conventional educational practices--for example, study of a single topic being confined within a given week of a course. The current results indicate that this compression of learning into a too-short period is likely to produce misleadingly high levels of immediate mastery that will not survive the passage of substantial periods of time (Cepeda et al., 2006, p.1101).

The research of Cepeda, Pashler, and their colleagues has produced a number of other results that are, at once, theoretically surprising and educationally significant. One of these concerns overlearning--the continuation of study immediately after the learner has demonstrated perfect, errorless performance on a given learning task. Recall from the paired-associates (somo -friend) experiment described earlier that, in the initial study session, participants were taken to a learning criterion of not one but two error-free recalls of all 40 word pairs. This modest degree of overlearning is a common design feature in the cognitive literature; some experiments enforce even higher levels, this ensuring that the to-be-learned material is indeed well learned. Overlearning is also a common memorization strategy among students at all levels (high school, university, etc.), especially in tasks that rely heavily [place a premium] on rote recall (foreign vocabulary drills, for instance).

But is overlearning an effective strategy? As recently as 2005, the answer appeared to be an unequivocal "yes" based on numerous reports of poorer retention under conditions of adequate learning (quitting practice once errorless performance has been demonstrated) versus overlearning (continuing practice past the point of perfection with no delay).

Today, however, a more nuanced view is called for. According to Rohrer and Pashler (2007), overlearning may be the only option when the failure to perform has dire consequences (which is why airplane pilots repeatedly practice how to handle in-flight emergencies). Overlearning is also an effective strategy in more prosaic cases where a person is content to retain information for just a short period of time, such as when a student spends the morning cramming for an afternoon exam. However, when the goal is long-term retention, measured in weeks, months, or years, then overlearning "simply provides very little bang for the buck" (Rohrer & Pashler 2007, p.184), because its benefits diminish rapidly over time. This doesn't mean that people can afford to skip practicing altogether and still ensure their long-term retention. Practice may indeed make perfect, particularly if it is distributed (or spaced) over several sessions than if the same amount of practice is concentrated (or massed) into a single session (Rohrer, Taylor, Pashler, Wixted, & Cepeda, 2005).

Overlearning is one of several old, well established cognitive concepts that are getting close new looks. Drawing in part on decades of research on operant conditioning, B.F. Skinner (1968) held that immediate feedback is essential for effective learning: instructors must not delay in letting learners know if a given response is correct , and instantly correct any response that is wrong. However, a more intricate and interesting account is suggested by Pashler, Cepeda, Wixted, and Rohrer (2005), who taught undergraduates obscure facts and tested their retention two weeks later. The results showed that when participants correctly recalled an item during initial learning, providing or withholding feedback had no appreciable effect on long-term (two week) retention. However, when errors occurred during learning, providing feedback with the correct answers boosted performance on the final test by 500%, relative to a no-feedback control condition. What is more, performance on the two-week test was significantly better when corrective feedback was given one day, rather than one second, after the commission of an error during learning. Though this remarkable observation is completely at odds with the Skinnerian dictum, it is a lovely illustration of what Bjork (1994) has termed <u>desirable difficulties</u>: conditions that impede, rather than promote, the progress of initial learning can enhance long-term retention and transfer of knowledge. We will revisit the concept of desirable difficulties later in this essay.

Spacing and Induction

Spacing is certainly a friend to the long-term recall of novel facts, word meanings, and other types of atomic items or associations. But is spacing an enemy of induction, or learning from examples? The question was recently raised by Kornell and Bjork (2008a), who pointed out that:

> In many everyday and educational contexts …, what is important to learn and remember transcends specific episodes, instances, and examples. Instead, it is important to learn the principles, patterns, and concepts that can be abstracted from related episodes or examples. In short, educators often want to optimize the induction of concepts and patterns, and there are reasons to think that such induction may be enhanced by massing, rather than by spacing (Kornell & Bjork, 2008a, p.585).

To test this idea, Kornell and Bjork carried out a series of studies in which undergraduates learned the styles of 12 different artists by viewing six different paintings (all landscapes or

skyscapes) by each artist. In one study (Experiment 1a), every participant studied six of the artists under conditions of massed presentation; for instance, the student was shown, in succession, six paintings by Georges Braque, followed by six consecutive paintings by Philip Juras, followed by six consecutive works by Marilyn Mylrea, etc. These same participants studied the other six artists under conditions of spaced presentation, so that paintings of any one of these artists were intermingled with the paintings of the other five; for instance, the sequence of presentation might be: Georges Seurat, Henri-Edmond Cross, Judy Hawkings, George Wexler, Bruno Pessani, a second Seurat, a second Wexler, etc.

 After the learning phase, students viewed four new paintings by the same 12 artists and tried to select, from a list of all the artists' names, the artist who had produced each new painting. Feedback was provided following every selection and wrong choices were always corrected with the right names. The selection task was divided into for blocks in order to evenly distribute the four choice paintings over all test trials. After the name-selection task, participants were told the meanings of the terms massed and spaced and asked: "Which do you think helped you learn more, massed or spaced?"
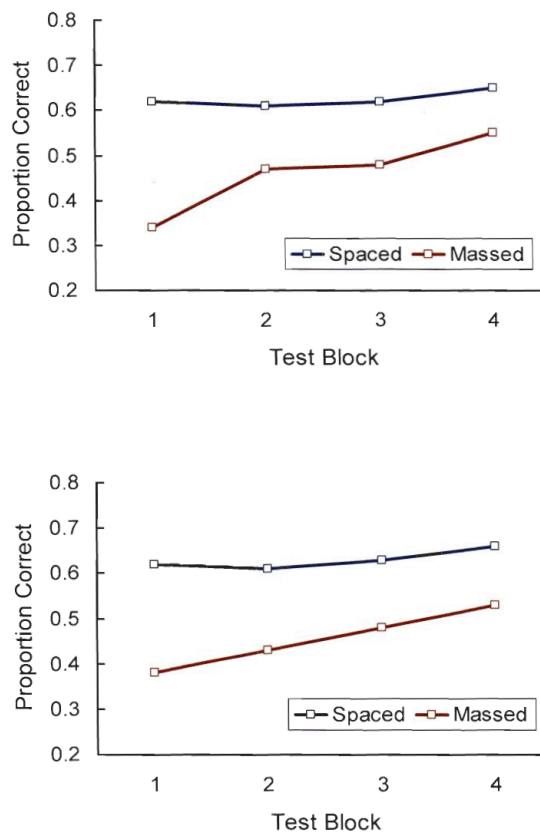

Figure 2: Mean proportion of artists selected correctly on the name-selection test as a function of test block and presentation condition. Source: N. Kornell & R. A. Bjork (2008a), Psychological Science, 19, 585-592, Experiments 1a (top panel) and 1b (bottom panel). Used with permission.

 Figure 2 shows the mean proportion of artists selected correctly on the name-selection test as a function of test block and presentation condition. Analysis of the data revealed a statistically significant improvement in performance over test blocks--a predictable result, given that corrective feedback was provided. However, in clear contrast to Kornell and Bjork's

hypothesis, performance on the name-selection test--a proxy for inductive learning--was reliably better under conditions of spaced as opposed to massed presentation in every block of trials. This startling result is no fluke: as shown in the bottom panel of Figure 2, the same pattern emerged in a second study (Experiment 1b) in which presentation condition was varied between rather than within subjects (such that every participant was exposed only to massed or only to spaced presentation trials, rather than some of each).
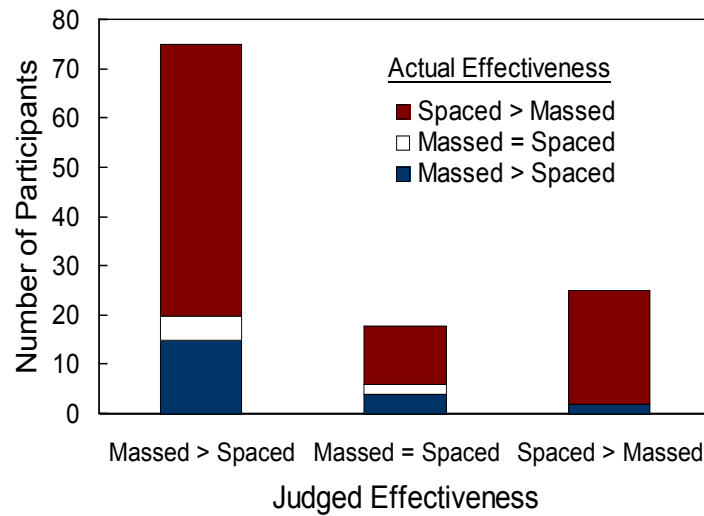

Figure 3 : Number of participants in Experiment 1a (out of 120) who judged massed presentation as more effective than, equally effective to, or less effective than spaced presentation. For each judgement, the number of participants is divided according to their actual performance in the spaced condtion. Source: N. Kornell & R. A. Bjork (2008a), Psychological Science, 19, 585-592. Used with permission.

The advantage of spacing over massing in inductive learning seems even more surprising in view of Figure 3, which summarizes the responses made by participants on the post-test questionnaire. As is apparent in the figure, most participants believed they learned more from massed than from spaced presentations, even though their performance on the name-selection test proved otherwise. In a curious coincidence of numbers, the percentage of participants-- 78%--that learned better with spacing than with massing was identical to the percentage who said that massing was as least as good as spacing.

Why didn't the participants' beliefs about learning effectiveness better reflect reality? Kornell and Bjork suggest that massing produces a sense of familiarity or fluency with the to-be-learned material that seduces subjects into thinking that they know the material better than they really do. That is, learners presume that the excellent short-term retention that massing all but guarantees will translate into excellent long-term retention, when in fact their feelings of familiarly or fluency are as fleeting as their memories. This and other types of cognitive illusions are significant road-blocks not only to learning and memory, but also to perception, comprehension, and transfer of training (see Koriat, 1997; Metcalfe, 2009).

As to why spacing promoted induction, Kornell and Bjork suggested several possible answers. One of these relates to the well-established finding, discussed earlier, that spacing strengthens memory for novel facts, word meanings, and other types of atomic items or

associations--a category that presumably includes artists' names, which were the targets of the retention tests used in Kornell and Bjork's studies. Had participants in these studies been asked to remember different styles of paintings, rather than the name of the artist associated with each style, an advantage of massed over spaced presentation might have emerged.

To test this idea, Kornell and Bjork ran a new study (Experiment 2) in which the learning phase was identical to the learning phase of their first study (Experiment 1a, described above). During the test phase, however, participants were asked to look at a previously unseen painting and decide whether it had been painted by a "familiar artist"--that is, a painter whose works had been presented during the learning phase--or an "unfamiliar artist." Even though this test placed a premium on remembering studied artists' styles, and not their names, the results produced a near-perfect replication of the earlier findings: a strong advantage in inductive learning of spaced over massed presentation and a striking discontinuity between actual test performance and the participants' metacognitive judgments about how they learn.

Another approach to understanding the interplay between spacing and induction goes back to early experiments (e.g., Kurtz & Hovland, 1956), in which participants learned to verbally discriminate between different categories of drawings; for instance, all drawings containing blue-tinted, rectangular objects were to be called "Kems" whereas drawings of roundish, greenish objects became "Javs." Inductive learning was superior when drawings from the same category were massed together, rather than interspersed with drawings from different categories--the opposite of Kornell and Bjork's results on the induction of artists' names or their styles of painting.

The latter investigators suggest that the difference may depend on how readily the instances of one category can be discriminated from those of another. When discrimination is straightforward, as it was in Kurtz and Hovland's (1956) research with simple line drawings, then massing may be advantageous to inductive learning. However, when discrimination is difficult, as in Kornell and Bjork's work with complex artistic stimuli, the advantage may go to spacing.

These ideas are part of a long list of issues involving spacing and induction that remain to be worked out and explored. The simple fact is that very little relevant research has been carried out to date, for the simple reason that, prior to Kornell and Bjork (2008a), it seemed so intuitively improbable that spacing could befriend both recall <u>and</u> induction. Apparently, cognitive psychologists are prone to a unique type of cognitive illusion that is ripe for dispelling. As Kornell and Bjork (2008a, p.591) point out:

> Inductive learning--that is, learning from examples--is a key element of formal education, and of how humans (and other animals) informally learn about the world. There are many inductive-learning situations that would seem, from an intuitive standpoint, to lend themselves to massed study, but may not. Examples include a baby leaning what <u>chair</u> means by observing people talking about chairs; an older child learning the rules of a language, such as that most plural English words end in <u>s</u>, by listening to people speak the language; a student in school learning how words are spelled by reading them (as well as through more direct instruction); a quarterback learning to recognize a complex pattern of motion that predicts an interception by gaining experience in practice and during games; a monkey learning to recognize the warning signs that another monkey is acting threateningly by observing other monkeys' behavior; and a medical student learning to recognize warning signs of lung cancer by reading x-rays under an expert's supervision. Our results cannot necessarily be generalized to all of these situations, of course, but they do suggest that in inductive-learning situations,

spacing may often be more effective than massing, even when intuition suggests the opposite.

Test-Enhanced Learning

Achievement tests are the Rodney Dangerfield of academe: they get no respect. As Roediger, McDaniel, and McDermott (2006) have remarked, rare is the student who relishes taking tests or teacher who enjoys giving them, especially when testing takes away valuable class time that could be put to better uses, such as instruction, discussion, and creative activities. In addition, many middle- and high-school teachers have serious doubts about the merits of standardized testing and strongly object to the practice of "teaching to the test"--concerns that are widely shared among parents, school administrators, and politicians. And cognitive psychologists have long looked upon educational tests (if they looked at all) as instruments that merely measure what students have learned, not as a means of altering the availability and accessibility of the students' knowledge (Marsh, Roediger, Bjork, & Bjork, 2007; Sternberg & Grigorenko, 2001).

Today a very different view of testing is gaining acceptance among cognitive scientists. It now appears that "frequent classroom testing (and student self-testing) can greatly improve education from kindergarten though university" (Roediger, McDaniel, & McDermott, 2006, p.28). Moreover, tests have been shown to improve long-term retention more than additional study of the material, even when tests are given without feedback.

To clarify, consider a two-session experiment by Roediger and Karpicke (2006a). During the first session, university students in group SSSS studied a prose passage covering general scientific concepts over four successive periods, each lasting 5 minutes. In contrast, participants in group SSST had three consecutive periods to study the passage, and then took a 5 minute test of free recall, writing down as much as they could remember in any order. Members of group STTT studied the passage just once, and then took three consecutive tests of free recall. Each study or test period lasted 5 minutes, meaning that the entire first session lasted 20 minutes for every participant in every group.

The second session took place either 5 minutes or 7 days after the conclusion of the first session. The participants' only task in this second session was to spend 5 minutes recalling as much as they could about the target passage.
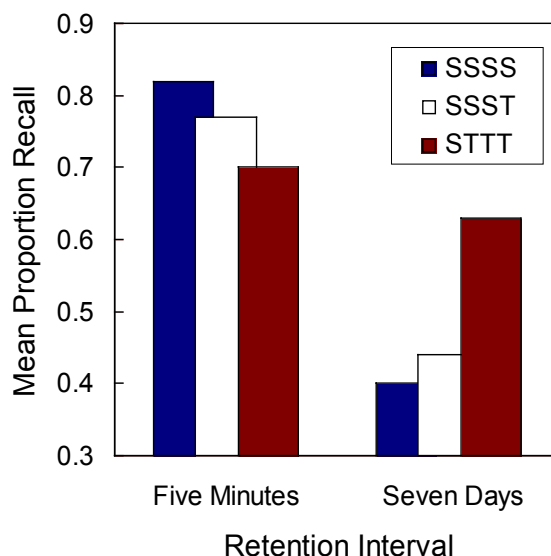
Figure 4 : Mean proportion of idea units recalled on the final test after a 5-minute or 7-day retention interval as a function of learning condition (SSSS, SSST, or STTT). The labels for the learning conditions indicate the order of study (S) and test (T) perdiods. Source: H.L. Roediger & J. D. Karpicke (2006a), <u>Psychological Science</u>, <u>17</u>, 249-255. Used with permission.

Figure 4 shows the mean proportion of "idea units"--key elements of the passage--that were recalled on the final test. When the retention interval was short (viz. a gap of 5 minutes between first and second sessions), performance on the final recall test tracked the number of study periods: the SSSS group clearly outperformed their STTT counterparts, while participants in the SSST group occupied the middle ground. However, when the retention interval stretched to a week, the pattern reversed: now the students who had studied the least, but been tested the most (group STTT), fared the best in final recall, while students who had devoted the entire first session to memorization (group SSSS) performed the worst.

These results, among many others (e.g., Karpicke & Roediger, 2007, 2008), demonstrate that testing is a potent tool for improving long-term retention. This holds not only for relatively simple tests that tap knowledge of basic facts and definitions, but also for tests of more complex capabilities, such as reasoning and understanding (Marsh et al., 2007). Moreover, the advantages of testing are apparent with many test formats, including free recall, short answer, and true-false questions (Roediger & Karpicke, 2006b), as well as with both open- and closed-book exams (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008). Interestingly, formats that place a premium on the active generation or production of information--short-answer exams and essays, for example--seem to produce greater improvements than do multiple-choice tests, which require the recognition of a correct answer against a background of alternatives or lures. As pointed out by Roediger and Karpicke (2006b), the benefits to long-term retention of producing (rather than recognizing) information at the time of testing mirror the benefits of producing (rather than simply reading) information at the time of study (the "generation effect" established by Slamecka & Graf, 1978). Put another way, people tend to learn best when they take an active role in both the encoding and the retrieval of to-be-learned material.

In practice, testing is clearly conducive to long-term retention. In principle, however, this boost in performance could backfire. According to Roediger and Karpicke (2006b, p.203):

> Although teachers would never deliberately provide false or misleading information during class or in reading materials, they routinely do so on some of the most popular kinds of tests that they give: multiple-choice and true/false tests. Many multiple-choice tests present three erroneous answers along with one correct answer. For some items, students may pick the wrong alternative, and because the act of retrieval enhances later retrieval, they may acquire erroneous information. Similarly, in true/false tests, typically half the items are true and half are false. Students may sometimes endorse false items as being true and thereby learn erroneous information. However, even if they read a false item and know it is false, the mere act of reading the false statement may make it seem true at a later point in time.

The issue at stake here is the <u>negative suggestion effect</u>: the increased belief in erroneous information that students may acquire from multiple-choice, true/false, or other commonly used tests. Though the effect was established long ago (Remmers & Remmers, 1926), recent demonstrations of test-enhanced learning have revived interest in the unwelcome prospect of test-diminished learning. After reviewing the literature, Roediger and Karpicke (2006b) concluded that although students may indeed learn false facts from multiple-choice and

other types of tests, the positive effects of testing far outweigh this cost. Further, negative suggestion effects can be reduced either by giving students test feedback (Butler & Roediger, 2008) or by offering them a "don't know" option, with a penalty for selecting a wrong answer. Marsh et al. (2007) have shown that this "free responding" strategy yields a small but significant decrease in lure production on a later test of cued recall.

One other point concerning test-enhanced learning merits comment, as it bears on the use of flashcards--a popular memorization technique among students of all ages ("Flashcard," 2009). As Kornell and Bjork (2008b) have remarked, studying with flashcards makes good sense, given that it involves both spaced practice and memory testing--activities that promote efficient learning and lasting retention. At the same time, most students assume that once they can correctly answer the question posed on one side of a flashcard (e.g., "What's the chemical symbol for silver"), without looking at the response printed overleaf ("Ag"), they can safely set that item aside and "drop" it from the deck.

On first impression, this strategy also makes good sense, inasmuch as it creates more opportunities to study the undropped items. However, in a series of four experiments, Kornell and Bjork (2008b, pp.133-134) found that:

> … participants did not profit from being allowed to self-regulate their study time by dropping items. If anything, dropping resulted in a small but consistent disadvantage. The disadvantage was not significant in every analysis, nor was it large in numerical terms, but it is truly surprising because there is a compelling reason to expect the opposite: Dropping ostensibly known items allowed participants to focus more study time on items they did not know. The average student would find the idea of spending equal time on all information when studying--even information they feel they already know--very foolish indeed.

In actuality, the average student would be wise to continue practicing known items in order to reap the benefits of test-enhanced learning. At the same time, continued practice ensures that learning trials are maximally spaced rather than increasingly massed--an unfortunate but necessary consequence of dropping the easiest-to-recall items first, and the most difficult items last. Another argument against the dropping strategy is that it can lead students to make ill-considered "deals" with themselves about how much time they should spend studying. In Kornell and Bjork's (2008b, p.134) experience, students sometimes "drop cards not to allow more time for others, but rather to hasten the end of the study session, because they refuse to stop studying until they have dropped all their cards." Such metacognitive decisions about how, what, and when to study are keys to the success, or failure, of self-regulated learning (Kornell & Bjork, 2007; Son & Metcalfe, 2000).

Applications and Conclusions

The research reviewed in the preceding sections represents a new--and long-overdue--trend among cognitive scientists to tackle educationally meaningful questions about learning, remembering, and forgetting with educationally meaningful materials, tasks, and retention intervals. Nonetheless, the fact that this research was conducted under tightly controlled laboratory conditions raises another educationally meaningful question: do the results generalize to more realistic learning environments (elementary school classrooms, university lecture halls, Web-based tutorial systems, etc.)?

To date, only a handful of translational projects been undertaken (see Pashler, Bain, Bottge, et al., 2007; Richland, Linn, & Bjork, 2007; Roediger, Agarwal, Kang, & Marsh, 2010), but the results are encouraging. Here are some examples:

- In a study by Shana Carpenter and her colleagues (Carpenter, Pashler, Cepeda, & Alvarez, 2007; also see Pashler et al., 2007), 8th graders reviewed facts learned in a U.S. history course either 1 week after learning them or 16 weeks later. On a final test given at the end of the 9-month course, students in the 16-week-delay group outperformed their 1-week delay counterparts by nearly 100%--a powerful and practical demonstration of the spacing effect.

- Mark McDaniel and his associates (McDaniel, Anderson, Derbish, & Morrisette, 2007; McDaniel, Roediger, & McDermott, 2007) developed an online brain and behavior course for University of New Mexico undergraduates. Every seventh day, the students either took a practice quiz on the course material they had covered that week or, as an exposure control, they were given additional time to read this same material. The quizzes contained either multiple-choice or short-answer questions, and students were given instant feedback on their performance.

  In addition to the practice quizzes, students received three criterial tests consisting of a pair of tri-weekly unit exams and a single cumulative final. On each of these tests, students scored significantly higher on questions pertaining to quizzed in contrast to non-quizzed material--evidence of test-enhanced learning. Importantly, this advantage was specific to quizzing--additional reading had no appreciable effect on criterial-test performance. Moreover, the magnitude of improvement was greater for short-answer than for multiple-choice quizzes, presumably because the former require more elaborative cognitive processing (Roediger et al., 2010). The benefits of quizzing seen in McDaniel's online course have been observed in other educational environments, including traditional college lecture halls and middle-school classrooms (Roediger, McDaniel, McDermott, & Agarwal, in preparation), as well as with a variety of educational materials, including scientific concepts and foreign language vocabulary (Roediger et al., 2010).

- Arguably the most ambitious translational project to date has been carried out by a research team led by Janet Metcalfe (2006; Metcalfe & Kornell, 2007; Metcalfe, Son, & Kornell, 2007). They compared the level of learning achieved following a computer-based study program with performance in two control conditions, self-study and no study.

  Participants in the initial investigation were 6th and 7th graders at a South Bronx school who were at high risk for academic failure. The to-be-learned materials were science terms and English vocabulary items, selected in consultation with the children's teachers. As described by Metcalfe (2006, p.27), the computer program was deliberately "overdesigned to include as many learning-enhancing principles from cognitive science as possible." These principles included elaborative processing, multimodal (sight plus sound) presentation, and self-generation of responses, along with several of the key concepts we covered earlier--spaced practice, corrective feedback, and test-enhanced learning.
  Every child participated in 5 weekly training sessions, switching between computer-based and self-study conditions every 25 minutes. The left panel of Figure 5

shows the results of a final vocabulary test given on Week 6. Relative to the no-study control condition, performance was nearly 9 times higher following the computer study program, but only about 2 times higher following an equal amount of time spent self-studying. The benefits of the cognitive-science based program were also evident in two follow-up studies: one in which Spanish-speaking children in the South Bronx learned English vocabulary (Figure 5, middle panel), the other in which English-speaking undergraduates at Columbia University learned Spanish vocabulary (Figure 5, right panel).
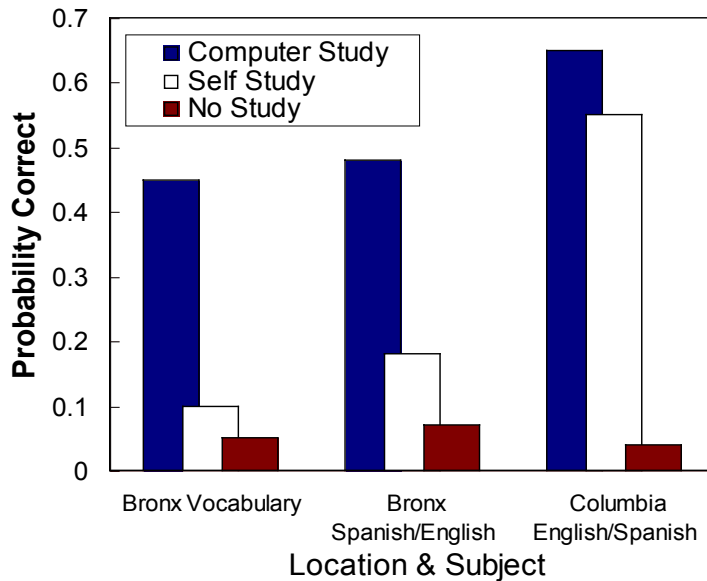

Figure 5: Mean performance on a final vocabulary test as a function of study condition for three groups of participants: South Bronx elementary students learning science terms and English vocabulary (left panel), South Bronx elementary school learning Spanish/English translations (middle panel), and Columbia University undergraduates learning English/Spanish translations (right panel). Source: J. Metcalfe (2006), <u>APS Observer</u>, <u>19(3)</u>, 27. Used with permission.

The studies by Carpenter, McDaniel, Metcalfe and others have made a valuable start toward bridging the gap between laboratory and classroom--but only a start. A host of issues and problems remained to be worked out and explored in future translational research. One key issue is how students would react to a curriculum that places a premium on spaced rather than massed practice, frequent testing instead of repeated studying, or other learning strategies that embody Bjork's (1994) concept of "desirable difficulties."

On the upside, Leeming (2002) developed special sections of two summer-semester courses (Introductory Psychology and Learning & Memory) in which students received an exam every day for 22-24 days--a stark change from his standard practice of giving 4 exams in each course, spread over the term. Final retention was measured after 6 weeks, and as one would expect, the results revealed test-enhanced learning: the exam-a-day students outperformed their 4-exam counterparts by a margin of 80% to 74% in the Introductory Psychology course, and by 89% to 81% in Learning & Memory. Moreover, on a post-course survey, the exam-a-day students reported studying more and being more interested in the course material. The

observations of Leeming and other instructors may help mollify critics who "wonder if college students would not rebel in shock at the introduction of weekly or even daily quizzing" (Roediger & Karpicke, 2006b, p.205).

Still, as Roediger et al. (2010, pp.41-42) have remarked:

> It is clear that when left to their own devices, many students engage in suboptimal study behavior. Even though college students might be expected to be expert learners (given their many years of schooling and experience preparing for exams), they often labor in vain (e.g., rereading the text) instead of employing strategies that contribute to robust learning and retention. Self-testing may be unappealing to many students because of the greater effort required compared to rereading, but this difficulty during learning turns out to be beneficial for long-term performance (Bjork, 1994). Therefore, the challenge for future research is to uncover conditions that encourage learners to set aside their naïve intuitions when studying and opt for strategies that yield lasting results.

Similarly, the challenge for educational institutions is to encourage instructors to set aside their naïve intuitions when teaching and opt for strategies that yield lasting results for their students. UBC is well positioned to meet this challenge: it already has many programs in place to help lecturers, faculty, and graduate students become better educators (e.g., the Centre for Teaching and Academic Growth), a commitment to original research on learning and retention (e.g., the Carl Wieman Science Education Initiative), mechanisms to recognize and reward excellence in teaching and research (e.g., Celebrate Learning and Celebrate Research weeks), and a culture that promotes novel approaches to instruction (e.g., the teaching innovations mentioned at the beginning of this essay). With these pieces in place, and with minds open to the possibility the long-term gains in the retention and utilization of knowledge may well require short-term pains during its acquisition, UBC is poised to become a leader in the cognitive science of learning enhancement.

Response to "The Cognitive Science of Learning Enhancement: Optimizing Long-Term Retention"

Professor Anthony Clarke, Department of Curriculum and Pedagogy anthony.clarke@ubc.ca   bio

First, let me say how much I enjoyed reading and learning from the paper written by Professor Eric Eich. It is clearly written, cogently argued, and well presented. It moves from definitions to examples to concepts and in doing so deeply engages the reader in the ideas being presented. The paper provides possibilities and conjectures about learning that makes one stop and think about one's own practices and examine a little more closely why it is that we believe in those practices.

Eric has successfully brought together a body of literature that spans many years but has long been in the background or ignored in most considerations of teaching and learning. His paper shines a light on two important ideas for the long-term retention of knowledge: 'massed and distributed learning' and 'test-enhanced learning.' Importantly, Eric presents these two ideas and the supporting literature in a balanced and reasoned fashion highlighting the strengths and

weaknesses associated with each. Additionally, he suggested conditions and contexts in which each idea might be best applied, eschewing the temptation for generalization without discrimination.

My response to Eric's paper is in support of and an extension to the work of Eric. I believe he correctly differentiates between the learning of content material that requires rote memorization (or is largely algorithmic in nature) and the learning of content material that requires a more conceptual appreciation (or is more relational in nature). The reader can quickly discern from the examples what and how the reported outcomes might relate their own teaching area.

Importantly, Eric's paper made me think more deeply about how I learn something for long-term retention—forcing me to 'go meta-cognitive.' I was particularly struck by his notion of "cognitive illusions", that is, the difference between how I think I best learn something and how I actually learn something for long-term retention. Using one of Eric's examples, I might think that the best way to learn a topic is to spend a long time reading and thinking about that particular topic and then at the end of that time put my ideas onto paper. However, the literature that Eric presents suggests that interspersing my reading and thinking with a regular 'self-test of knowledge' might actually increase my long-term retention that information. An interesting idea and one that Eric suggests we might consider introducing into appropriate classroom contexts.

While Eric's paper prompted me to think more deeply about what I do as an instructor and it also prompted me to think more deeply about who are the students in my classroom. As I was thought about how I might improve their long-term retention of knowledge, I couldn't help but wonder about the dispositions, the prior knowledge, the assumptions about my subject area, and their motivation for learning that they bring to my classroom. An understanding of these attributes is important to what and how they learn in the same way that the learning environment that I provide is important to what and how they learn. Thus, it is important to bring the two into relation with each other.

I recommend Eric's paper to all those interested in how we might promote learning in university classrooms be they clinical one-on-one settings or large-group lectures. There is something for all educators to think about in this paper.

References

Agarwal, P.K., Karpicke, J.D., Kang, S.H.K., Roediger, H.L., & McDermott, K.B. (2008). Examining the testing effect with open- and closed-book tests. Applied Cognitive Psychology, 22, 861-876.

Anderson, J.R. (2007). How can the human mind occur in the physical universe? New York: Oxford University Press.

Anderson, R.C., & Faust, G.W. (1973). Educational psychology: The science of instruction and learning. New York: Dodd, Mead.

Baddeley, A. (1990). Human memory: Theory and practice. Boston: Allyn and Bacon.

Beal, C.R., Qu, L., & Lee, H. (2008). Mathematics motivation and achievement as predictors of high school students' guessing and help-seeking with instructional software. Journal of Computer Assisted Learning, 24, 507-514.

Beck, C.D.O., & Rankin, C.H. (1997). Long-term habituation is produced by distributed training at long ISIs and not by massed training or short ISIs in Caenorhabditis elegans. Animal Learning & Behavior, 25, 446-457.

Bjork, R.A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), Practical aspects of memory II (pp.396-401). London: Wiley.

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

Bottge, B., Rueda, E., & Skivington, M. (2006). Situating math instruction in rich problem-solving contexts: Effects on adolescents with challenging behaviors. <u>Behavioral Disorders</u>, <u>31</u>, 394-407.

Bransford, J.D., & Donovan, S.M. (2005<u>). How students learn: History, mathematics, and science in the classroom</u>. Washington, DC: National Academy Press.

Butler, A.C., & Roediger, H.L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. <u>Memory & Cognition</u>, <u>36</u>, 604-616.

Carpenter, S. K., Pashler, H., Cepeda, N.J., & Alvarez, D. (2007). Applying the principles of testing and spacing to classroom learning. In D.S. McNamara & J.G. Trafton (Eds.), <u>Proceedings of the 29th Annual Cognitive Science Society</u> (p.19). Nashville, TN: Cognitive Science Society.

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

Cepeda, N.J., Coburn, N, Rohrer, D., Wixted, J.T., Mozer, M.C., & Pashler, H, (2009). Optimizing distributed practice: Theoretical analysis and practical implications. <u>Experimental Psychology</u>, <u>56</u>, 236-246.

Cepeda, N.J., Vul, E., Rohrer, D., Wixted, J.T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. <u>Psychological Science</u>, <u>19</u>, 1095-1102.

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

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

"Flashcard." (2009, October 13). In Wikipedia, the free encyclopedia. Retrieved October 19, 2009, from http://en.wikipedia.org/wiki/Flashcard

<u>Goettl, B.P. (1996). The spacing effect in aircraft recognition. Human Factors, 38, 34-49.</u>

Heckler, A.F., Kaminski, J.A., & Sloutsky, V.M. (2008). Learning associations that run counter to biases in learning: overcoming overshadowing and learned inattention. In B.C. Love, K. McRae, & V. M. Sloutsky (Eds.), <u>Proceedings of the 30th Annual Conference of the Cognitive Science Society</u> (pp. 511–516). Austin, TX: Cognitive Science Society.

Karpicke, J.D., & Roediger, H.L. (2007). Repeated retrieval during learning is the key to long-term retention. <u>Journal of Memory and Language</u>, <u>57</u>, 151-162.

Karpicke, J.D., & Roediger, H.L. (2008). The critical importance of retrieval for learning. <u>Science</u>, <u>319</u>, 966-968.

Klahr, D., & Li, J. (2005). Cognitive research and elementary science instruction: From the laboratory, to the classroom, and back. <u>Journal of Science Education and Technology</u>, <u>14</u>, 217-238.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. <u>Journal of Experimental Psychology: General</u>, <u>126</u>, 349-370.

Kornell, N., Bjork, R.A. (2007). The promise and perils of self-regulated study. <u>Psychonomic Bulletin & Review</u>, <u>6</u>, 219-224.

Kornell, N., & Bjork, R.A. (2008a). Learning concepts and categories: Is spacing the "enemy of induction"? <u>Psychological Science</u>, <u>19</u>, 585-592.

Kornell, N., & Bjork, R.A. (2008b). Optimizing self-regulated study: The benefits and costs of dropping flashcards. <u>Memory</u>, <u>16</u>, 125-136.

Kurtz, K.H., & Hovland, C.I. (1956). Concept learning with differing sequences of instances. <u>Journal of Experimental Psychology</u>, <u>51</u>, 239-243.

Leeming, F.C. (2002). The exam-a-day procedure improves performance in psychology classes. Teaching of Psychology, 29, 210-212.

Lesgold, A., & Glaser, R. Eds. (1989). Foundations for a psychology of education. Hillsdale, NJ: Erlbaum.

Marsh, E.J., Roediger, H.L., Bjork, R.A. & Bjork, E.L. (2007). The memorial consequences of multiple-choice testing. Psychonomic Bulletin & Review, 14, 194-199.

McDaniel, M.A., Anderson, J.L., Derbish, M.H., & Morrisette, N. (2007). Testing the testing effect in the classroom. European Journal of Cognitive Psychology, 19, 494-513.

McDaniel, M.A., Roediger, H.L., & McDermott, K.B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. Psychonomic Bulletin & Review, 14, 200-206.

Metcalfe, J. (2006). Principles of cognitive science in education. APS Observer, 19, 27 & 38.

Metcalfe, J. (2009). Metacognitive judgments and control of study. Current Directions in Psychological Science, 18, 159-163.

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, 530–542.

Metcalfe, J., & Kornell, N. (2007). Principles of cognitive science in education: The effects of generation, errors and feedback. Psychonomic Bulletin and Review, 14, 225-229.

Metcalfe, J., Kornell, N., & Son, L.K. (2007). A cognitive-science based program to enhance study efficacy in a high and low-risk setting. European Journal of Cognitive Psychology, 19, 743-768.

Pashler, H. (2006). How we learn. APS Observer, 19, 24-26.

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). Organizing instruction and study to improve student learning (NCER 2007-2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ncer.ed.gov

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

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

Pasnak, R., Cooke, W.D., & Hendricks, C. (2006). Enhancing academic performance by strengthening class-inclusion reasoning. Journal of Psychology: Interdisciplinary and Applied, 140, 603–613.

Remmers, H.H., & Remmers, E.M. (1926). The negative suggestion effect on true-false examination questions. Journal of Educational Psychology, 17, 52-56.

Richland, L.E., Holyoak, K.J., & Stigler, J.W. (2004). Analogy generation in eighth grade mathematics classrooms. Cognition and Instruction, 22, 37–60.

Richland, L.E., Linn, M.C., & Bjork, R.A. (2007). Cognition and instruction: Bridging laboratory and classroom settings. In F. Durso (Ed.), Handbook of applied cognition: Second edition (pp.555-583). New York: Wiley.

Roediger, H.L., Agarwal, P.K., Kang, S H.K, & Marsh, E.J. (2010). Benefits of testing memory: Best practices and boundary conditions. In G.M. Davies & D.B. Wright (Eds.), New frontiers in applied memory (pp.13-49). Brighton, UK: Psychology Press.

Roediger, H.L., & Karpicke, J.D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. Psychological Science, 17, 249-255.

Roediger, H.L. & Karpicke, J.D. (2006b). The power of testing memory: Basic research and implications for educational practice. Perspectives on Psychological Science, 1, 181-210.

Roediger, H.L., McDaniel, M., & McDermott, K. (2006). Test-enhanced learning. APS Observer, 19, 28.

Roediger, H.L., McDaniel, M.A., McDermott, K.B., & Agarwal, P.K. (in preparation). <u>Test-enhanced learning in the classroom</u>.

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

Rohrer, D. & Pashler, H. (2007). Increasing retention without increasing study time. <u>Current Directions in Psychological Science</u>, <u>16</u>, 183-186.

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. <u>Applied Cognitive Psychology</u>, <u>20</u>, 1209-1234.

Rose, J.K., Kaun, K.R., & Rankin, C.H. (2002). A new group-training procedure for habituation demonstrates that presynaptic glutamate release contributes to long-term memory in <u>Caenorhabditis elegans</u>. <u>Learning & Memory</u>, <u>9</u>, 130-137.

Siegler, R.S., & Ramani, G.B. (2006). Early development of estimation skills. <u>APS Observer</u>, <u>19</u>, 34-44.

Skinner, B.F. (1968). <u>The technology of teaching</u>. New York: Appleton-Century-Crofts.

Slamecka, N.J., & Graf, P. (1978). The generation effect : Delineation of a phenomenon. <u>Journal of Experimental Psychology : Human Learning and Memory</u>, <u>4</u>, 592-604.

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

Sternberg, R.J., & Grigorenko, E.L. (2001). All testing is dynamic testing. <u>Issues in Education</u>, <u>7</u>, 137-170.