How Students Learn

-- and How We Can Help Them

John F. Kihlstrom

Department of Psychology

University of California, Berkeley

Note: This paper is based on a presentation to the Working Group on How Students Learn sponsored by the GSI Teaching and Resource Center, University of California, Berkeley, March 8, 2011. It incorporates new material presented at the Teaching Conference for GSIs on January 18, 2013, and at a seminar on How Students Learn on March 6, 2013 -- both sponsored by the GSI Teaching and Resource Center. I thank Linda von Hoene for the invitation to develop this presentation, and Judith M. Harackiewicz for comments on the section concerning intrinsic motivation and interest.

See also "Teaching and Technology", invited presentation to the Academic Senate, University of California, Berkeley, April 23, 2013. Link to written version.

If we want to know "How students Learn", it is perhaps good to begin at the beginning, with a definition of learning -- something that
psychologists have been studying since the late 19th century, when Pavlov rang his first bell and Thorndike put his first cat in a puzzle box. While under the yoke of Watsonian and Skinnerian behaviorism, psychologists defined learning as "a relatively permanent change in behavior that occurs as a result of experience". After the cognitive revolution, the definition got revised in one important way: we now think of learning as a relatively permanent change in knowledge that occurs as a result of experience. That knowledge is reflected in the organism's behavior, but the important thing is that learning changes the individual's fund of knowledge.

Learning and Memory

That knowledge may or may not translate into behavior, but by virtue of learning it becomes available for use, stored in memory. Now, psychologists distinguish among a number of different kinds of memory, including "short term" or "working" memory and "long-term" memory. When we talk about student learning, we're mostly talking about "long-term" memory -- though working memory is not by any means irrelevant. Distractions, like music or checking your cell phone for texts, can consume some of the capacity of working memory, with the result that relatively little can be encoded in or retrieved from long-term memory.

Cognitive psychologists commonly distinguish among various types of knowledge stored in long-term memory. There is, first, a distinction between declarative and procedural knowledge. Declarative knowledge is factual knowledge, about what is true or false, which can be represented in sentence-like structures known as propositions. Procedural knowledge is knowledge of skills and rules, how to do things, which can be represented in "if-then" conditional statements known as productions.

- There are two kinds of procedural knowledge: motor procedures, like how to shift gears in a standard-shift car, and mental procedures, like how to calculate square roots.
- And there are two kinds of declarative knowledge: episodic knowledge is essentially autobiographical memory, for particular events that have a unique location in space and time; semantic knowledge is more abstract, like a mental dictionary or encyclopedia. The names of the planets are recorded in my semantic memory; the fact that I learned them for a science project in Mrs. Sly's third-grade class is recorded in my episodic memory.

 Much of what we know about memory we know about episodic memory, and then generalize to the other types. But there's one special fact about procedural knowledge, which is enshrined in what Anders Ericsson, of Florida State University, has called the 10,000-Hour Rule (made famous by Malcolm Gladwell in his best-
selling book, Outliers) -- it takes about 10,000 hours of practice to get really good at a skill. That's 40 hours per week, 50 weeks a year, for 5 years; or 3 hours a week, 7 days a week, for 10 years; and it's the difference between Yo-Yo Ma and the rest of us. By that standard, most students probably need to study harder.

Psychologists have been studying memory for about 125 years, ever since Ebbinghaus invented the nonsense syllable, and we know quite a lot about how it works -- enough so that cognitive neuroscientists can begin to figure out how the brain does it. But I'm not going to talk about the brain today. There is enough to say about the learning process at the psychological level of analysis.

Principles of Memory

We begin by distinguishing between three stages of memory processing -- encoding, the process by which a new trace is laid down in memory; storage, or what happens to the encoded memory trace over the retention interval; and retrieval, or gaining access to stored knowledge so that you can use it to solve problems or whatever. The fate of memory over each of these three stages is governed by a remarkably small number of principles.

The encoding stage involves, two principles, elaboration and organization. Ebbinghaus' original theory of memory, based on British associationism, was that memory was fixed by rehearsal -- by simply repeating the item to be remembered, over and over again. But we now know that maintenance rehearsal, just repeating something to ourselves, over and over again, like we would a name or telephone number, is not sufficient to encode that item in long-term memory -- which is why, if we're interrupted, the thing we're rehearsing goes right down the mental drain. Instead, what's needed is what is known as elaborative rehearsal, connecting up what we're trying to learn with what we already know. I haven't talked with my old travel agent, in Madison, Wisconsin, for more than 20 years, but I still know his phone number -- 256-4444, because when he gave it to me he pointed out that 256 = 4 raised to the 4th power. He only had to say it once. Examples like this give us the elaboration principle, that memory is best when we process an item deeply, connecting it with our rich fund of pre-existing knowledge. So here's the first key to learning: we learn best when we learn progressively, building new knowledge on old knowledge.

Actually, almost anything that encourages the reader to pay close attention to a text will improve memory. A recent study by Connor Diemand-Yauman and his colleagues at Princeton (2010) showed that presenting text in an unfamiliar font, such as Comic Sans, which is relatively hard to read, led to better memory for text contents compared to a more familiar font, such as Arial (this website is mostly formatted in
Verdana). The more effort you expend, the better you'll remember -- provided that it's not just rote rehearsal.

The elaboration principle is supported by a further organization principle: memory is best when we relate the things we are trying to learn to each other, to see how they are connected together, or share certain features. If you group items into categories, or find some other links among them, you can learn much faster than if you study each item on its own. The elaboration principle deals with "item-specific" processing, while the organizational principle has to do with "inter-item" processing. But the essential principle is the same: we learn things best when we attend to how they relate to other things.

The storage stage is governed by the time-dependency principle, which also has its roots in Ebbinghaus -- not to mention your grandmother: memory gets worse over time. But why does memory get worse over time. Memories might fade, the way a photograph does; or they might be kicked out by newly arriving memories, as when a filing cabinet gets filled. Both these principles apply to some forms of memory: if you don't engage in active processing, items will drop out of working memory; and the capacity of working memory is limited to something like the famous "seven, plus or minus two" items. But so far as long-term memory is concerned, we think that once encoded, knowledge is permanently stored in memory, and the problem is to get the memory out. What causes forgetting from long-term memory, then, is neither decay or displacement, but rather interference -- other bits of stored knowledge get in the way. In fact, there's a "paradox of interference" in which the more you know about a subject, the harder it is to retrieve any particular piece of information about it. But the paradox can be overcome if you've organized the material, so you can retrieve it efficiently.

Encoding makes knowledge available in memory storage, but retrieval allows us to gain access to stored information. On the retrieval side, perhaps the most important principle is cue-dependency -- the accessibility of a memory is a function of the informational value of the cue used to retrieve it. That's why multiple-choice tests are easier than short-answer or fill-in-the-blanks tests: multiple-choice tests are, essentially, recognition tests whose cues are a virtual copy of the information stored in memory. So, if you want to know what a student has really learned, short-essay, short-answer or fill-in-the-blanks tests are probably better than multiple-choice tests. But they're also less efficient and less reliable in the scoring, so there's a tradeoff here that has to be considered: You can cover a lot more material with a multiple-choice test, and you can grade the tests much more reliably and efficiently.

The principle of cue-dependency is qualified, to some extent, by a principle of encoding specificity -- memory is best when the cues used to retrieve information match those that were processed at the time the memory was encoded. To take an example from my introductory psychology class, if you encode AMBER as a mineral, you're going to find it difficult to retrieve that word as a girl's first name. So, again, when you study you want to make as many different
connections to the material as possible, to ensure that a wide variety of cues will effectively retrieve it when you want it.

Memory is better when the context in which retrieval is attempted matches the context in which it's encoded. It's really true that students do better on exams administered in the same classrooms where they've taken the course. But this context dependency must be qualified by cue-dependency. Sometimes the cues contained in the query to memory overshadow the cues in the environmental context, so that subjects are more likely to show context dependency on short-answer questions, as opposed to multiple-choice questions.

The context is internal as well as external, and sometimes subjects think that, if they've studied with a beer (or a joint) in hand, they should take the test that way, too. But no: alcohol, marijuana, and similar psychoactive drugs do induce "state-dependency" in memory, but they also exert a general negative effect on both encoding and retrieval processes. If you study with a buzz on, you may score better on a test you also take with a buzz on, compared to if you were sober, but there is no question that study goes better, and so does test-taking, if you're sober at both times.

But there is an interesting twist on encoding specificity, which is that how subjects expect to be tested affects how they will perform on test day. Subjects who study anticipating a free-recall test do better if they actually take a free-recall test, compared to a recognition test, and vice-versa. Apparently, we encode memories differently, depending on how we expect to retrieve them. So, again, students should be encouraged to study for a wide variety of tests, so that their encoding strategies don't disadvantage them when the crunch comes.

Retrieval and encoding aren't entirely separate and independent: they interact with each other in various ways. For example, there is a compensatory relationship between encoding and retrieval: lots of retrieval cues (as in a multiple-choice test) may make up for poor ("shallow") processing at the time of encoding; but by the same token, "deep" processing at the time of encoding can help performance even on difficult tests that don't supply many cues.

Another way that encoding and retrieval interact is through a principle of "use it or lose it". One of the implications of the elaboration principle is that frequent testing improves memory. Every time we retrieve a memory we encode it again, laying down a new memory trace next to the old one (as it were), but each time a little different. These different versions of the same memory, created through repeated testing, make it possible to retrieve information quickly when we need it again later. The lack of repeated testing is what accounts for the "summer slump" that besets students (and their teachers), especially in math and science, between spring term and fall term. You've got to use this knowledge or you'll lose it, and one way to insure that you'll use it is to arrange for frequent testing. The same principle applies within a semester, as between semesters. It doesn't do students any good to go through an entire term, and then administer a comprehensive final
at the end. Frequent testing, including repeated testing of the same material, helps them retain information better.

The relations between encoding and retrieval are also illustrated by the principle of **schematic processing**: memory for a piece of information depends on the relation of that information to pre-existing knowledge (organized in the form of a cognitive structure known as a *schema*). Information that is congruent with our prevailing schema is remembered better than information that is irrelevant to it: this is because the schema supplies cue information that is useful at the time of retrieval, thus illustrating the principle of cue-dependency. But, interestingly, information that is incongruent with our prevailing schema is remembered best of all: this is because schema-incongruent events violate our expectations, and demand explanation, and this explanatory activity creates a richer, deeper encoding. The point here is that the cognitive background is critical for learning. So learning proceeds best when the student already has some background, on which he or she can build, and which can provide a framework for generating expectations and questions.

One illustration of the schematic processing effect is a somewhat paradoxical **testing effect**, by which testing someone *before* they attempt to learn material enhances their memory for that material after learning -- even if they fail to get the answer right on pretesting. Prior testing allows the individual to get some sense of the kinds of things he's going to be tested on. Moreover, simply *trying* to find the answer in memory appears to enhance the encoding of that answer, once it's encountered in the study session. Although it's a pain, for both instructor and students, a good argument can be made for testing students' knowledge of course material before the course begins. Not only does that enable the instructor to assess what a student has actually learned in the course, but the very act of preliminary testing appears to enhance learning and memory.

All of the principles discussed so far illustrate what we call the **library metaphor** of memory. Pieces of knowledge are like books, which must be acquired, cataloged, and placed on shelves, where they can be checked out and read. The library metaphor will take you a long way, but in the end it's not quite right. It turns out that the information stored in memory traces is typically vague and ambiguous, and needs to be filled out with inferences based on world-knowledge. A similar situation obtains in perception, where the stimulus is typically vague and fragmentary, and the perceiver must engage in constructive activity that fills in the gaps, and permits the subject to, in Jerome Bruner's immortal phrase, "go beyond the information given" in the stimulus. In memory, we refer to this as **reconstruction**, hence the **reconstruction principle** of memory: remembering is less like reading a book than writing one from your notes.

The reconstruction principle brings us full circle, back to elaboration. You've got to have some world knowledge in order to learn and remember. The same world-knowledge that fosters elaborative processing at the time of encoding, fosters rich reconstruction at the time of retrieval. And remember, the act of retrieval doesn't
just strengthen a single memory trace; it also encodes an entirely new trace, that in some sense sits alongside the old one.

I should also note that psychologists distinguish between two expressions of memory: explicit memory is conscious recollection, while implicit memory is the unconscious influence of some past event on the person's experience, thought, or action. And there are also two forms of learning: explicit learning is what we ordinarily mean by learning, while implicit learning is the kind of learning that takes place when you're not consciously aware of what you've learned. I take it that in this workshop we're more interested in conscious learning and memory, and that's what the seven principles just outlined are all about. But unconscious learning also takes place, more or less incidentally, in the ordinary course of everyday life. We learn all the time, without necessarily knowing what we've learned.

And implicit memory can actually be a boon to students at test time: on a multiple-choice test, for example, if we don't know the answer, one of the options may simply "ring a bell" -- an effect of implicit memory, in the absence of explicit memory. So there's a test-taking strategy that all students should know: if you know the answer, put it down and don't go back; if you don't know the answer, but can eliminate at least one option as clearly wrong, that will increase your chances of getting the question right. But if you don't know the answer, but one of the options seems familiar, choose it. Unless the instructor is fiendishly tricky, if we've done the work, attended class and done the readings, that choice will be right more often than wrong.

**PQ4R**

All of this laboratory research boils down to a practical system for studying and learning, called the **PQ4R method** by John Anderson, after a strategy originally formulated by Thomas and Robinson (1972). It's also called the SQ3R method, but it's the same idea.

- **Preview** -- Before reading, survey (that's where the S comes from in SQ3R) the material you're trying to learn. Look at the section headings, note any new terms that are in bold face. Get yourself ready to learn and remember by establishing an appropriate schema.
- **Query** -- Based on your survey, make a list of questions that you want answered by the text, so that you can read the text with those questions in mind.
- **Read** -- Read carefully. Don't skim -- you already did that in your preview. Read with the intent of answering the questions you formulated.

http://socrates.berkeley.edu/~khlstrm/GSI_2011.htm
• **Reflect** -- After you've read the chapter (or, better yet, a smaller section of that chapter), think about how what you've read relates to the questions you formulated (this is the fourth R, added by Anderson to the standard SQ3R). Go over any examples used. Think about how the material relates to what you already know.

• **Recite** -- Now put the text aside and try to recall the material that you've read by answering the questions you formulated earlier.

• **Review** -- Go back and compare your recall to what you read, and see if there are any discrepancies that need correction.

To which we can add a fifth R:

• **Repeat** -- Now go back and do it again, perhaps after reading the next chapter or section. Students tend to think that if they've read some material once, that's enough. It isn't. There's a reason that we say that they should spend at least two, maybe three, hours outside class for every hour they spend inside class. For a student taking 12 credits (assuming that one isn't volleyball, but even if it is), that means 24-36 hours of study per week outside of class.

Going to college is a full-time job, but students don't always realize this. Surveys tell us that the average college student spends only about 2-3 hours per night studying. If you do the math, that's not enough.

The PQ4R method places great weight on *questions* -- both asking them and answering them. In a seminal study by Frase (1975), a group of subjects read a text with an instruction to generate questions based on it. A second group were given the questions generated by the first group, and instructed to read the text with the goal of answering them. A third group just read the text. Later, all subjects were given a test on which some of the items were relevant to the text, and others were not. Subjects in the two groups did much better on this test than did those in the control group.

**The Importance of Teaching Earnestly**

So far, all of this is what any cognitive psychologist could tell you about how students learn. But, as my title indicates, there's more to the psychology of learning than cognitive psychology -- there's a social psychology, too. Which is why learning isn't just for cognitive psychologists anymore.

In fact, long ago, Neal Miller and John Dollard (1941) advanced the concept of
social learning theory to account for complex human behavior. They asserted that most learning is about social behavior, and most learning takes place in a social context. Stanford's Albert Bandura expanded on their view, distinguishing between various modes of learning:

- **Learning by Direct Experience** -- this is the usual "trial and error" learning, familiar from the studies of classical and instrumental conditioning that you studied in your introductory psychology course. Pavlov's dog hears a bell and then tastes food, and be pretty soon forms an association between the two. Thorndike's cats get put in a container, and pretty soon they learn to press a paddle in order to open a latch and escape. Pavlov, Thorndike, and practically everyone else who studied learning well into the 20th century, all thought that the animal was passive during conditioning -- which is where the term "conditioning" comes from. One exception was Berkeley's Edward C. Tolman, who argued that even animals were actively engaged in the learning process, trying to figure out what is going on in their environment, and what to do about it -- a position ridiculed by some who characterized Tolman's maze-running rats as "lost in thought at the choice point". But we now know that Tolman got it right. The learning organism is actively generating, testing, and revising expectations and hypotheses about its world; when it learns, it learns to predict and control events in its environment. That's true for dogs and cats, rats and pigeons -- and it's also true for humans.

- But Bandura was more interested in a different type of learning, one that had also been described by Miller and Dollard -- **observational learning**, or learning vicariously, not by direct experience, but by observing other people. We do learn by direct experience, but we also learn from other people, and learning from other people is by far the more efficient. If you stick your finger in an electrical socket, you'll eventually learn not to do that; but if I tell you not to do it, you'll learn a lot more quickly. It's observational learning, learning from other people, that lies at the heart of social learning theory.

Bandura further distinguished between two major forms of observational learning:

- **Learning by Example**, which is to say, by observing other people. This includes various forms of imitation, some of which occur automatically and unconsciously, as well as conscious deliberate, modeling.

- **Learning by Precept**, including both informal and formal, sponsored teaching. Many social interactions involve one person in the role of teacher and the other in the role of learner. And society has developed a wide variety of institutions to generate and conserve knowledge, and transmit it to the next generation -- not the least of which is the college and university. There is no more efficient way for students to learn, I think, than from a well-organized course of lectures accompanied by a well-written textbook.
Which brings us to the subject of teaching, and the obvious point that students learn by being taught. It follows from this, I think, that students learn best if they're taught well. And one of the keys to teaching well, the literature tells us, is for the teacher to have command of the material being taught, and to organize it in such a way that students are able to master it (e.g., Pascarella et al., 2008).

Arthur Graesser and his colleagues have produced an expanded list of "25 Learning Principles for Pedagogy and the Design of Learning Environments", accompanied by a concise overview of the supporting literature.

All of this is in the service of **mastery learning**. Ideally, students should not progress to later material in a course, unless they have truly mastered earlier material. This is because the earlier material forms the cognitive background -- the **schema** against which subsequent learning can take place.

Mastery learning is epitomized by the **Personalized System of Instruction** (PSI) developed by Fred Keller (1968), a friend and colleague of B.F. Skinner's, who taught for many years at Columbia University (and before that, at my alma mater, Colgate). The PSI may be the best thing to come out of the entire behaviorist tradition in psychology -- not least because it's not all that "behaviorist", despite having been published in the house-organ of Skinnerian behaviorism. But I digress.

Keller proposed that a course -- introductory psychology, say -- be divided into modules -- thematically coherent chunks that are smaller than a textbook chapter -- on which students work independently with access to a proctor who can answer questions and correct mistakes. Each module concludes with a mastery test, and the student is not permitted to advance to the next module unless and until s/he achieves a score of at least 90%. If he doesn't achieve this criterion, he returns to the module and tries again, and keeps trying until he meets this "unit perfection requirement".

Note that, given, the unit perfection requirement, every student who completes the course is guaranteed an A. Keller had no problem with this, and neither do I. The PSI entails **power** testing, not **timed** testing: the only purpose of the test is to insure that the student has mastered the material. Any student who masters at
least 90% of the material in a course deserves an A. You could give other letter
grades, I suppose, based on the number of cycles the student needed to achieve
the criterion of mastery, or you could loosen the criterion to allow students to
proceed after scoring only 80%, or 70%. But then you've voided the basic purpose
of the PSI, which is to insure that students have mastered the material. And
you've also violated the basic, if implicit, assumption of the PSI -- which is that
every student can master the material.

The PSI requires a high level of course structure, and imposing that structure is the
job of the teacher, but the teacher's role is performed in advance. The teacher
may offer lectures and demonstrations, but these are mostly for purposes of
motivating students, not providing information to them. The course content comes
from the modules, so that there is a great deal of emphasis on the written word.
Once the modules (and the tests) are constructed, the student works
independently, with access to the proctor who tutors and grades the tests -- and
also provides a little human contact, thus enhancing the social in social learning.

The "go at your own pace" format means that, in order to be properly
implemented, the PSI requires extended periods of time. It may take more than a
quarter or a semester for the student to achieve mastery of the material. And that
has to be OK. And because the PSI is personalized, it's hard to implement in a
classroom. But clever people can probably figure out a way to integrate the PSI
with ordinary classroom activities.

And there are reasons for thinking that people should work on that,
because the PSI is a highly effective system of instruction. Meta-
analyses of studies of the PSI, reviewed by Pascarella and Terenzini
(1991, 2005) show that the PSI is much more-effective than
traditional methods -- the difference amounting to "effect sizes"
of .40 or greater -- which are pretty substantial in social-science and educational
research. Something very much like the PSI is used in some lower-division
programming courses at Berkeley, and a version of the PSI is used at Carnegie-
Mellon University to teach the introductory psychology course. For those who are
interested, it's perfect for online instruction.

Keller entitled his paper on the PSI "Good-Bye, Teacher", but it's pretty clear that
the teacher is still pretty important. It's up to the teacher to determine the
material to be covered, divide the material into modules, present the modules
(whether in writing or in pre-recorded lectures) so that they get the material across
to the student, and create methods of evaluation that can be meaningfully repeated
if necessary. In some circumstances, the teacher may also be called upon to serve
as proctor. Like most technical innovations in education, whether it's PowerPoint of
the Internet, it requires more work from the teacher, not less -- especially when
compared to the ease of pulling out yellowed, decade-old lecture notes, or just
reading aloud from the textbook (which, I swear on my copy of William James's
Principles of Psychology, I have witnessed both an assistant professor do in intro
psych, and a tenured professor do to a graduate proseminar).
Learning Techniques and Instructional Strategies

There is an increasing, and increasingly rigorous, body of research on the effects of various learning techniques and instructional strategies, especially in the context of high-school and college courses, and especially in the context of the so-called STEM disciplines of science, technology, engineering, and mathematics (where, frankly, the criteria for achievement and mastery are more intuitively self-evident).

For example, a recent paper by John Dunlosky and his colleagues (2013) evaluated the utility of ten popular learning techniques, with surprising and somewhat disconcerting results. Two of the techniques most favored by students -- re-reading and highlighting their texts -- proved pretty useless in terms of such educational outcomes as test score and long-term retention. And two of the most useful techniques proved to be those which are loathed by students and faculty alike -- namely, practice testing and distributed testing. teachers don't want to be constantly constructing, administering, grading, and discussing tests; and students don't want to be constantly taking them, either. And teachers want to get through their syllabus, instead of returning to material covered previously in the course (much less prerequisites for the current course). Yet those very techniques are the most useful in promoting student learning.

The most-effective learning techniques aren't just aggravating -- they're also counterintuitive. Another recent paper, by Rohrer and Pashler (2010), focused on three such points.

Testing is for Learning

In the PSI, as in conventional teaching, the primary purpose of testing is to assess, or demonstrate, the student's mastery of the material. But it's now also clear that testing is critical to the learning process itself -- we learn by being tested, or by testing ourselves. Testing improves retention, even more than extra study, and this is true even when subjects don't get any feedback from the test. This is called the testing effect (Roediger & Karpicke, 2006), and it's as true in applied educational settings as it is in the sterile confines of the psychological laboratory. Repeated testing helps, even when students get questions on the practice tests wrong.
In a study by Roediger and Karpicke (2006), students were asked to read a typical text. After a two-minute delay, they either read the text again (the study-study condition), or were given a written recall test of what they had read (the study-test condition). They then received a final written recall test after a delay of 5 minutes, 2 days, or 1 week. Rereading helped performance, a little, after the shortest delay. But after the longer delays, those who had taken the test showed vastly superior scores.

A more recent experiment by Karpicke and Blunt (2011) at Purdue University also made this point dramatically. Subjects who read a passage and then were immediately tested on that passage retained about 50% of the passage a week later, compared with a control group who simply read the passage over and over -- what might be called cramming for a test. But interestingly, the testing group also outperformed a "concept mapping" group that drew diagrams to represent what they had learned. Both cramming and concept mapping increased student's estimates of what they had learned -- but these estimates were illusory. Subjects in the testing group thought they had learned less, but they actually learned more, as evidenced by their test performance a week later. Interestingly, the testing effect improved performance on direct questions about the contents of the text, and also questions about inferences drawn from the text, but not actually specified therein.

The testing effect has been replicated many times, with a wide variety of materials and in classroom as well as laboratory contexts. And it is not simply a result of the fact that the test provides additional exposure to the test material. That much is obvious from the comparison between the "retrieval practice" and "cramming" conditions of the Karpicke & Blunt experiment. Testing encourages deep, elaborative processing at the time of testing, when the material being tested is being re-encoded into memory. And the act of testing itself creates different pathways for retrieval, increasing the likelihood that students will find the knowledge in storage when they need it.

Massed vs. Distributed Practice

Speaking of cramming, another consistent finding of research is that study and practice sessions are best spaced out over relatively long intervals of time. This is known as the spacing effect: if the same amount of study time is distributed across several sessions, as opposed to being compressed into a single session, more learning gets accomplished, and it last longer.

This figure, a real thing of beauty, shows the magnitude of the spacing effect in one recent experiment (Cepeda et al., 2008). After the subjects studied a list of unfamiliar facts, they studied them again after an interval varying between 20 minutes and 105 days (you read
that right). Then, 7 to 350 days after the second study session (you read that right too), they were tested on their recall of the material. Scores were highest after relatively short retention intervals, but even after 350 days retention was best if at least a little time had elapsed between the two study sessions. If you really want to know, Cepeda et al. estimated that the optimal gap between study sessions is about 5-10% of the study-test delay. So, if you want students to remember on the final exam something that they read in the first week of class, they should read it again a week or two later. If you want someone to remember something for life, they should study the material again at least a year after their first exposure.

Like the testing effect, the spacing effect has been replicated with lots of different kinds of study materials, and in classroom as well as laboratory settings. The mechanism underlying the spacing effect isn't well understood, and probably has to do with how individual events are encoded in memory. But as an empirical fact it couldn't be clearer.

**Interleaving Problem Types**

Another surprising finding has to do with how things are presented for study. Consider a math or science course, where students are trying to learn how to solve problems of somewhat different types. It might be expected, from what I said earlier about mastery learning, that the best strategy would be to require students to master one type of problem before going on to another. But it turns out that this sort of **blocked presentation** isn't the best educational strategy. **Interleaved presentation** is better pedagogy.

Here's a sample study, by Rohrer and Taylor (2007, Exp. 2), in which subjects learned to find the volumes of fairly obscure geometrical solids, like wedges, spheroids, spherical cones, and half cones. some subjects worked on the problems in a blocked fashion (aaabbbccccdd), while for other subjects the problems were interspersed (e.g., abcdabdcacadb). Interleaving Spacing made initial learning a little more difficult, but improved performance on a test administered 1 week after the learning trials.

This **interleaving effect** has been replicated with lots of different kinds of study materials, and in classroom as well as laboratory settings. It's not just a special case of the spacing effect, because the effects of interleaving persist even when the spacing between problems is equated between the blocked and interleaved conditions. The underlying mechanism has to do with how individual events are encoded in memory. But as an empirical fact it couldn't be clearer. And especially in mathematics, where both textbooks and homework assignments typically involve blocked practice, the implications couldn't be clearer either.
The Structure of Tests

Taken together, these three effects and the spacing effect argue not just for multiple midterm exams, and cumulative final exams, but also for testing material covered by the first midterm again on the second midterm. The testing effect is strongest when the repeated tests are spaced out over an extended period of time, rather than massed together. And, in fact, an "expanding" schedule is best -- with the first test immediately after studying some material, then a second test a little while later, a third test after a somewhat longer interval, and the like. So, to return to the PSI framework, we wouldn't want to test students just once, at the end of a module, and then never again. You'd want a schedule that cycles back to earlier material. Of course, with all this testing, there might not be time for anything else in the classroom. But at the very least, where a textbook comes with a study guide, students should be encouraged to make use of the practice tests -- not just to memorize answers, but to actually enhance their learning with repeated retrieval.

A particularly interesting form of testing is dynamic testing, in which students receive immediate feedback after each test item, instead of at the end of the test as a whole -- feedback that doesn't just tell them whether they're right or wrong, but also helps them to understand the principles involved, and where they went wrong. A related idea is formative assessment (as opposed to merely "summative" assessment). In formative assessment, we do more than simply tote up scores on tests and return them to the students. Instead, teacher and students alike treat the fate of individual test items as feedback which will guide both the teacher's teaching and the student's learning. When tests are viewed as an intrinsic part of the learning process, instead of an after-the-fact assessment of learning outcomes, tests enhance learning. That doesn't make them any less annoying -- for teachers to make, or for students to take; but at least they have a function.

Let me say something here about the issue of "teaching to the test". Especially in this age of high-stakes testing, and No Child Left Behind, and its analogous programs in higher education, there have been a lot of complaints about this practice. I have never understood this complaint. If the test is a good test, the test should be a reliable guide for both the teacher and the learner. I went to high school in New York State in the Sixties (the salad days of Averill Harriman and Nelson Rockefeller), and if you were going to get the "Regents" diploma required for college admission you had to pass statewide exams, set by the Board of Regents of the University of the State of New York in various subjects, including literature and history as well as science and mathematics. In fact, you didn't have to take the courses themselves. All you had to do was pass the test. And so our teachers taught to those tests, and we studied to those tests. But this was all right, because the tests were good tests -- they were valid assessments of the student's knowledge of some
subject. If you wanted to pass the test, you actually had to learn some chemistry, or history, or English literature, or whatever.

And what made them good tests was not just that they had the usual psychometric properties that we expect of any test -- standardization, norms, and reliability. They also were valid tests -- you had a sense that in various ways, they accurately measured the trait -- knowledge of chemistry or history or literature that they were supposed to measure. There was a symbiotic relationship between curriculum and test -- the curriculum didn't just reflect the test, which is what people complain about when they complain about "teaching to the test"; the test also reflected the curriculum.

In an incredibly stupid cost-saving move, in 2011 Regents voted to abandon foreign-language examinations. But here was another case where the exam actually led to a very positive restructuring of the curriculum. When I was in high school, in the 1960s, the foreign-language exams (I took them in Latin and German) were pretty plain-vanilla -- vocabulary and translation. But in the 1980s, the Regents exams changed: you had to show an ability to converse in the language, and you had to know something about the culture(s) where the language was spoken. And the curriculum changed accordingly -- precisely because teachers were teaching to the test.

Things turned around, though, when New York adopted the Common core standards, which 45 states had chosen to structure their primary and secondary curricula in mathematics and English and language arts (in the future, Common Core standards will be developed for other areas of study). The first examinations based on the Common Core standards were administered in 2013, and aroused much angst among students, teachers, and parents, because the tests covered material that they hadn't been taught, along with dire predictions of low scores and consequences for upper-school placement ("A Tough New Test Spurs Protest and Tears" by Javier C. Hernandez & Al Baker, New York Times 04/19/2013). But of course, that's the point: the tests represent the curriculum that is to be taught. If that's "teaching to the test", so be it.

So how can we construct valid tests. Psychometricians distinguish among various types of validity.

- **Face Validity** is the extent to which test items seem valid "on their face", or superficially, as measures of the trait (a test of extraversion shouldn't be composed of items relating to neuroticism). This is the simplest form of validity, and it simply means that a math test should contain math problems (though, of course, they can be word problems) and a history test should contain questions about history).

- **Content Validity** is the extent to which the test surveys the universe of
content represented by the trait. So, in introductory psychology, a proper test would have questions about personality as well as about learning theory, and questions about perception as well as about memory. When I create tests in my courses, I try hard to make sure that there's at least one concept tested from each lecture, and from each major section of the textbook chapter. You can't test every concept, but at least you can draw a representative sample.

- **Empirical Validity** is the extent to which test scores correlate positively with some criterion. This is, perhaps the hardest kind of validity to establish because we simply don't record that kind of information. When I pass students from my introductory course into Art Shimamura’s course on human learning and memory, it would be nice to know that students’ scores on my tests -- at least when I test their knowledge of the psychology of learning and memory -- correlate with performance on his tests. But, as I say, we don't keep records of that sort. We probably should, but we don’t.

- But at the very least we can guarantee that our tests have a sort of **Internal Consistency**. This is, technically, an aspect of reliability, not validity, but you'll see why I talk about it here. That is, scores on various test items should correlate with each other, so that the test as a whole "hangs together". In my classes, I edit my exams retrospectively, to eliminate items that don't correlate with performance on the test as a whole. Doing so insures that each item on the test really belongs there, and obviates most questions about whether an item (or a test) was "fair".

The point of all of this is that testing is aggravating for everyone, students and teachers alike. I like to lecture, even when I’m not particularly good at it, but I hate to make up tests. But testing, well done, is well worth the time and effort -- because testing itself enhances learning.

**Learning Styles**

The PSI demands a lot of the teacher, but it also demands a lot of the student, too, and such a mechanical system, which allows students to progress at their own pace but which otherwise is basically a "one size fits all" mode of instruction, may not be suitable for all students. Here I am thinking of the widespread attention that modern educational theory and practice gives to the notion of **learning styles** -- that students differ widely in terms of how they learn, and that the most effective teaching is tailored to each individual student's needs. In principle, at least, these individual differences in learning style are independent of general intelligence. They represent how a person applies his intelligence, not how much intelligence he has.

This very attractive idea is very old, having its origins in the theory of **psychological types** proposed by C.G. Jung -- this was after his falling-out with Freud. Jung believed that individual differences in
personality could be construed within an eightfold typology created by crossing two attitudes with four functions. The attitudes constitute the person's orientation to the external objective world or the internal, subjective world. The four functions are sensing versus intuiting, and thinking versus feeling. There's also a distinction between perceiving and judging. These differences are measured by the Myers-Briggs Type Indicator, a personality questionnaire very popular in both educational and business-management circles.

Another familiar scheme, proposed by Riley Gardner and his colleagues, has its origins in "neo-Freudian" psychoanalytic ego psychology, which tried to connect psychoanalysis with the experimental psychology of learning, perception, and memory. Gardner's differences had to do with such cognitive dimensions as leveling-sharpening (whether the person emphasizes or de-emphasizes differences), and field-dependence or -independence (the extent to which perception of an object is influenced by its surrounding context).

Perhaps the most popular approach to learning styles comes from the work of David Kolb (1984), based on what he called the cycle of learning. In Kolb's view, the learner begins with immediate concrete experience; then, based on observation and reflection, develops a theory and makes deductive inferences from it; these inferences are then tested against new concrete experiences, and the cycle begins again.

We all go through such a cycle, Kolb believed, but we differ in terms of which aspects we prefer, or are particularly good at. The four aspects of the learning cycle generate two bipolar dimensions, which in turn generate four types of learner. The "accommodator" emphasizes concrete experience and active experimentation, while the "assimilator" emphasizes reflective observation and abstract conceptualization. Kolb's framework was adapted by a consulting firm, the Hay Group, with different labels. Their "decision-maker" emphasizes abstract conceptualization and active experimentation, while their creator" emphasizes reflective observation and concrete experience. But it's the same idea, and their assessments are based on Kolb's instruments.

One more example, perhaps the most elaborate of the lot, is the "Building Excellence" program of Rundle and Dunn, which combines perceptual and psychological styles like visual-word vs. visual-picture and global vs. analytic along with physiological variables like
preferred time of day, environmental features like temperature and lighting, emotional qualities like conformity, and sociological characteristics like a preference for working alone vs. in large groups. The resulting scheme is depicted in this very striking figure which -- when you look at it -- is impossible to realize in three dimensions. But it does capture the essential features of what Harold Pashler and his colleagues (2008) have called the learning styles hypothesis.

It is important, though, to remember just what "the learning styles hypothesis" states. It doesn't just mean that individuals have different study preferences. There's no question that there are such individual differences in preferences -- as anyone who ever had a roommate knows. But reliable study preferences do not necessarily imply the existence of actual differences in learning style. The learning style hypothesis is that individualizing instruction to match the individual's learning style will maximize learning outcomes. Or, put another way, that teaching should "mesh" with the student's learning style.

As appealing as this hypothesis is, it is all the more surprising to learn that it has never actually been successfully tested. Pashler et al. (2008) reviewed the extant literature on learning styles, and found not a single study that unambiguously supported the learning-style hypothesis. In fact, they found only a single study that even ambiguously supported the learning-style hypothesis. Maybe such evidence exists in the proprietary data banks of some educational consulting firm. But it's not in the journals.

It's important to understand just what kind of evidence would support the learning-styles hypothesis. It comes in the form of what are known as aptitude by treatment interactions (ATI) in which the investigator compares the effectiveness of two different instructional methods (say, verbal vs. pictorial) with two different types of students (say, those with verbal and visual thinking styles), resulting in four (2x2, or what is known in psychology as the "Noah's Ark" design) experimental conditions. Plotting the average results in each of the four conditions, the learning-styles hypothesis is supported by what is known as a cross-over interaction -- such that, for example, outcomes are better when verbal learners are given a printed text than when they watch a video, and when visual learners watch a video instead of being given a printed text. There are many varieties of ATI, but they all have to have this cross-over feature; and, to make a long story short, none of the published research yields such an interaction.

Here's the single possible exception to the rule, from Robert Sternberg and his colleagues (1999). Sternberg is a distinguished cognitive psychologist who has devoted his entire career to the study of intelligence, the development of new and better intelligence tests, and putting educational practice on a firm evidential base. As a result of his research, he has identified three quite different types of intelligence, which he characterizes as analytic, practical, and creative. He took a group of high-school students taking introductory psychology in summer school, measured
them on these three types of intelligence, and then randomly assigned them to three different versions of the course -- one emphasizing abstract analysis, one emphasizing practical application, and one emphasizing creative problem-solving. And then he gave them a series of tests of their analytic, practical, and creative thinking about what they had learned. At first glance, the results seemed to support "analytic students taught analytically did better on analytic questions; practical students taught pragmatically did better on pragmatic questions; and creative students taught creatively did better on creative questions. But that's not the same as analytic students taught analytically doing better over all than analytic students taught creatively, and creative students taught creatively doing better over all than creative students taught analytically. Moreover, even these results involved quite a bit of data massaging and selective analyses (about which, I should say, Sternberg and his colleagues were quite up front).

And that's it. As I say, it's possible that the evidence supporting the learning-styles hypothesis may exist somewhere, in the files of some educational consulting firm. But the evidence is not to be found in the peer-reviewed literature on educational psychology.

The learning-styles hypothesis is very appealing: Who could be against tailoring education to maximize individual outcomes? But even if we had the evidence, implementing the hypothesis could become a very expensive proposition. We'd have to identify learning styles at the outset. Then, we'd have to train teachers to teach to all the different learning styles. But while the teacher is spending time with the visual learners, what are the verbal learners supposed to do? Or are we going to have different teachers specialize in teaching to one learning style or another, and assign students to different classrooms based on admissions testing? What happens when these kids get to college, where the teachers are barely trained to teach to one style? Or in the workplace, where there's only one manager, or supervisor, and they're not trained to teach at all? Maybe this can all be done, but it better be cost-effective. That is to say, the cost of implementing such a program better be worth the results it delivers. If it turns out that everybody learns OK by text and lecture, but verbal learners do a little more OK than visual learners -- that would be a crossover ATI, but the gains might not be worth it as a matter of social policy.

The other thing that should be said is that, viewed uncritically, the idea of accommodating individual differences in learning style can quickly become a prescription for discrimination. I'm thinking of claims by certain psychologists that men and women think "in a different voice", and other essentialist arguments for "difference feminism", which could lead to the segregation of boys from girls, and men from women, in schools and colleges. And I'm also thinking of Milwaukee's experiment with "immersion schools", in which African-American boys are segregated from everybody else, and taught mostly, by African-American male teachers, according to an "Afro-centric" approach. Maybe "separate but equal" is the right educational policy, but we had better be sure.
Procrastination

On a somewhat lighter note, one characteristic that students seem to have in common is a tendency toward procrastination (unlike their teachers, of course). Given a deadline, they (not to mention we) find it all too easy to put off the work until later. It's a real problem, and not just in the classroom, as James Surowiecki made clear in his recent New Yorker article ("Later: What Does Procrastination Tell Us About Ourselves?", 10/11/2010). There's actually a cognitive explanation for this, known as hyperbolic discounting. Put briefly, it turns out that the future looks farther away than it really is. In the meantime, short-term considerations tend to overwhelm long-term goals. Instead of working on a paper due at the end of the semester, a student might well go out to the movies on Wednesday night, or a basketball game. And do the same thing the next week, and the next, until it's crunch time and he's pulling an all-nighter trying to read Moby Dick in order to compare it to Ulysses (which he also hasn't started yet) -- or in your office pleading for an extension. Which, if you give it to him, just puts him farther behind on next semester's work.

It turns out that students actually prefer firm deadlines, as Dan Ariely, a psychologist who is one of the leading figures in behavioral economics, discovered when he asked them. And teachers can help them by imposing them. Instead of simply saying that a term paper is due at the end of the semester, require an outline at the end of the fourth week, and a draft at the end of the eighth week, and the final version at the end of the twelfth week (so you don't spend intersession grading papers -- a task which you, for your part, will put off as long as you can).

Parenthetically, I have to say that Berkeley's adoption of a one-week "Reading, Recitation and Review" period, replacing our old two-day "Dead Days" between the end of classes and final exams, is not necessarily in the best interests of either students or faculty. When I taught at Harvard, in the late 1970s, classes for Fall Term ended before Christmas, and then students returned to campus in January for a reading period and exams. The academic calendar has been changed now, after considerable resistance and debate. But until that reform, there was apparently an institutionalized form of behavior known as "to Harvard" (as in, "I'm going to Harvard this assignment"). After classes, some students would hole up in a hotel room, reading and writing under a sunlamp, and then return to Cambridge with a fine tan, amazing their students with the fact that they could vacation over the holidays and still get their work done. (The story may be apocryphal, and I don't know where I got it, but it might have been Scott Fitzgerald -- but the image makes my point).

Anyway, Ariely and Wertenbroch (2002) did a series of studies on academic procrastination, which suggest a number of techniques that will help students
overcome it. First, as noted earlier, they found that students actually prefer it if instructors establish firm, progressive deadlines for completing papers and other assignments. Second, even in the absence of self-imposed deadlines, students can be encouraged to set their own -- but when they do this on their own, the deadlines they impose on themselves are suboptimal -- too far away from the present, too close to the final deadline. So they need externally imposed deadlines that force them to spread their work out.

The key to meeting deadlines, then, is what Ariely and others call "extended will" -- external constraints on task completion. Students should be given focused tasks rather than open-ended tasks; the focused tasks should be broken up into manageable segments; and segment deadlines should be close, not distant.

Students can also help themselves by putting themselves on a reward schedule for meeting various deadlines. Instead of going to the movies on Wednesday night, go to the movies on Wednesday night if you've met the deadline for Tuesday. Buy yourself a cheeseburger once you've outlined Moby Dick. Buy yourself another one when you've figured out how to compare and contrast Ishmael with Leopold Bloom. Of course, all of this depends on the student's motivation to do any of this stuff, and his or her willingness to forego the movie tonight for the cheeseburger next week, which brings us back to the student determinants of learning.

Sleep

Not only do students procrastinate, but they don't get enough sleep -- or enough of the right kind of sleep. Adolescents and young adults are sleep-deprived almost by definition, especially if they try to sleep with their cell phones on; and nontraditional, returning students with heavy work and family obligations suffer threats to their sleep as well. And we know that sleep deprivation, especially Stage REM (where dreaming occurs), impairs the encoding of new memories, and probably memory retrieval as well. A sleep-deprived mind is just not a very efficient mind. Students who stay up all night cramming for an exam suffer a double whammy: By cramming, they rely on massed practice when spaced practice is better; and they are both studying and taking an exam when they're not thinking straight. I guess this counts as a triple-whammy.

If so, there's a fourth sleep-related whammy. My colleague Matt Walker has been at the forefront of researchers demonstrating that sleep -- particularly "slow-wave" or "Stage NREM" sleep -- plays an important role in consolidating memory, increasing its accessibility for later use. So at only do sleep-deprived students fail to encode new memories optimally; they also fail to optimally consolidate whatever it is they manage to encode (e.g., Walker & Stickgold, 2006).

There's only so much we can do about this, of course. But at least we can warn
students that they need a lot more sleep than they think they do, and a lot more sleep than they're getting. And if they would go to bed at a decent hour, and turn off their cell phones so that they don't hear the pings of incoming text-messages, their academic work would certainly profit (I thank Andrew Zeri, Dean of the Graduate Division at UC Berkeley, for reminding me of this point).

Implicit Theories of Learning

Classic theories of learning, of the sort that came out of the behaviorist tradition, hold that reward is critical for learning. Organisms, whether they're children or adults or pigeons or rats, learn if they're rewarded. This view was enshrined in two of Thorndike's classic laws of learning. According to the Law of Readiness, behavior was energized by various motivational states, such as hunger (or the desire for a "secondary" reinforcement, such as money). And according to the Law of Effect, learning occurred when the organism's behavior was reinforced by some event -- delivery of a food pellet, or a dollar -- that satisfied that motive.

But we now know that reward isn't necessary for learning to occur. Organisms, whether they're rats or humans, just learn from moving around in the world. This is what Tolman discovered in his famous studies of "latent learning" in rats. Another learning theorist, Harry Harlow (he of "motherless monkeys" fame), gave hungry rhesus monkeys a series of puzzles in which they had to figure out how to work a set of locks to open a door to get a treat (monkeys love Froot Loops). Given the Law of Readiness, the obvious control condition was one in which they monkeys weren't hungry. And given the Law of Effect, another obvious control condition was one in which the monkeys weren't rewarded with food. Harlow discovered that motive and reward didn't matter: the monkeys loved just sitting around learning to open the puzzles. It didn't matter whether they were rewarded; and if they weren't hungry, they squirreled the Froot Loops away for later eating.

Turning to the human case, the Canadian psychologist Douglas Berlyne argued that we come into the world with a great deal of epistemic curiosity. We want to know how the world works, and it doesn't matter whether we're rewarded for learning such things. Similarly, Arie Kruglanski talks about individual differences in need for closure: ambiguity makes most of us pretty uncomfortable, and we want to resolve it. Neither of these needs have anything to do with Froot Loops -- or money, either, for that matter. We come into this world ready, willing, and able to learn. And that's what we do -- something gets in the way.

One of the things that can get in the way is the individual's implicit theory of competence. In psychology, an implicit theory is like a scientific theory, but less formal, and less clearly articulated, and less subject to rigorous hypothesis-testing and revision. As Berkeley's own Allison Gopnik has argued so cogently, children
are in the business of developing theories about themselves and the world around them (she calls this the "theory theory" of development).

- Children develop implicit theories of the physical world as they learn to throw and catch balls.
- They develop implicit theories of biology as they discover the difference between living and nonliving things.
- They develop implicit theories of sex when they ask "Where did I come from", and their parents tell them about storks or "special hugs".
- They develop implicit theories of gender when they learn that some things are "for boys" and other things are "for girls".
- They develop implicit theories of personality when they observe the different kinds of children and adults they interact with.
- They develop implicit theories of society as they work out the kinship and status relations of the people they live with.
- They develop implicit theories of learning as they interact with schools and teachers (including those who teach them at home).
- And they develop implicit theories of intelligence and competence in much the same way.

And this process of theory development, testing, revision, and more testing, continues throughout the life cycle, from birth to death.

Carol Dweck and her colleagues (she's now at Stanford) have distinguished between two implicit theories of competence that, in her view, affect children's and adults, motivations to learn.

- In **entitativity**, the person believes that competence (intelligence, whatever) is a fixed quantity -- an entity. This fixed quantity may be innate (in the genes) or acquired (a product of learning), but once you've got it, you keep it; and if you don't get it, you're not going to get it. Students who are entity theorists tend to explain failure in terms of a lack of ability, and they stop trying. Teachers do the same thing. I remember well a colleague whose high-school daughter was having trouble in math. She approached her teacher, who watched her work on a sample problem, and remarked, "You just can't do this, can you?". Somewhat later, Lawrence Summers, then Harvard's president, explained the dearth of women on the tenured faculty of Harvard's math and science departments by invoking the possibility that, by virtue of their genetic endowment, females just don't have the chops for math (never mind that a similar gender imbalance can be found in Harvard's humanities and social-science departments).
- By contrast, **incrementalism** inclines toward a view of competence as malleable, something that can be changed through one's own efforts. If you work hard enough -- think of the 10,000-hour rule -- you can get good at something. And, in fact, students who are incrementalists explain failure in terms of their own lack of effort. You don't have to be Amy Chua, the infamous Tiger Mother, to understand that if students, and their teachers,
think they are capable of learning something new and difficult, they're more likely to do it.

Dweck and her colleagues have found that entity and incremental "theorists" are differently motivated with respect to earning. For entity theorists, the primary goal is to look smart, while for incremental theorists the primary goal is to actually learn something. And, somewhat paradoxically, praising children for being smart tends to undercut their efforts at school, and thus their performance -- for one thing, it leads them to want to avoid making mistakes. Or they think that they can rely on their raw intelligence to pull them through. Or even: those who are praised for being smart, but know they're really not, will just decide that there's no point to working hard in school. Better to praise the children's efforts: for Dweck, to say someone is "hard working" isn't to damn them with faint praise. And the concept of "overachiever" isn't viable, because it implies that the student is doing better than he or she is supposed to do, given his or her endowment of intellectual abilities. It's not important how smart someone is (or is supposed to be). What's important is how hard one works; and people who work hard may find out that they're smarter than they think they are.

This is important, because Dweck and her colleagues have also found that students who are induced to abandon entititativity, and adopt an incrementalist view instead, actually do better in their studies. They do this by describing the brain as a muscle that gets stronger when it's exercised -- and by teaching them some study skills (like PQ4R). That's because incremental theorists believe that they can do better if only they tried harder. Entity theorists don't think there's anything they can do - - the best they can do is cover up their insufficiencies.

Now, most of Dweck's research has been done on elementary and middle-school students, and application of her "growth mindset" principles at the college level doesn't always work out. But in one provocative study, Joshua Aronson and his colleagues enrolled Stanford students in a program in which they wrote "pen-pal" letters to local schoolchildren, basically touting the virtues of incrementalism. What was interesting was that the college students themselves began getting better grades, and enjoying their studies more. But things haven't always worked that way. Still, if one of the purposes of a college education is to promote intellectual development, it seems that it would help if both students and their teachers believed that such development was possible.

Dweck's work also suggests a possible correction to my earlier disparaging comments about learning styles. In a sense, entititativity and incrementalism are approaches to learning. Not only are there individual differences on this dimension, there are also cultural differences. A fairly substantial cross-cultural literature, for example, indicates that the "East Asian" cultures of Japan, Korea, and China tend to assume incrementalism: there the focus is on the value of working hard, trying your best, and striving for perfection. By contrast, Western European culture tends to lean toward entititativity -- as exemplified by its reliance on things like IQ tests to measure intellectual ability as a stable quantity. Actually, IQ is more malleable than
we think it is. But it's not a great leap to think that, at the same time as East Asian cultures are adopting American institutional structures for secondary and higher education, Western students could gain a lot from adopting the incrementalist viewpoint characteristic of the East.

**Intrinsic Motivation**

The work of Tolman and Harlow, Berlyne and Kruglanski, and a host of others indicates that we need no extrinsic motivation to learn. As behaving organisms, we're wired, innately to learn about the world we live in. And as conscious, sentient beings (we're *homo sapiens*, after all) we have both a lot to learn, and a powerful apparatus -- including the capacity for language and the culture that language and consciousness, operating together, create -- with which to learn it. But that doesn't mean that there aren't, as well extrinsic motives to learn as well. We learn not only because we want to learn, just for the sake and satisfaction of learning. We also want to learn because we're rewarded for learning, and because learning enables us to achieve other things that we want -- beginning with food and shelter and safety, but also status and self-esteem.

Psychologists have come to distinguish between two broad sources of motivation. **Extrinsic motivation** is a person's desire to engage in some specific activity in order to achieve some goal or satisfy some need -- in brief, to gain some reward. Some extrinsic rewards are **primary**, in that they meet some biological need. Other extrinsic rewards are **secondary**, in that they are more symbolic, and derived from primary-rewards. Money is a good example: you can use it to buy food. Monkeys will work for grapes, but they'll also work for poker chips with which they can purchase grapes. Academic grades are also good examples, because they stand for something else -- though, as we'll see, precisely what they stand for turns out to be important.

**Intrinsic motivation**, by contrast, refers to a person's desire to engage in some specific activity without any promise or prospect of reward. In the present context, intrinsic motivation refers to a student's intrinsic interest in schooling, and intrinsic desire to achieve competence in his or her studies.

At first glance, you would think that having both types of motivation going in the same direction would be really good. That is, if someone already wants to do something (like study), rewarding them for doing that thing should make them want to do it even more. But, as always, things are not that simple, and it turns out that sometimes extrinsic rewards can undermine intrinsic motivation, so that the person becomes less interested in that activity than he was before.

Among the first to notice this was Mark Lepper, a psychologist at Stanford University. In a classic experiment, Lepper and his

colleagues asked nursery-school children to do something that little kids like to do anyway -- draw on big sheets of paper with Magic Markers. Some of the children were promised a "Good Player" award if they did this; others; weren't promised anything, but unexpectedly got the "Good Player" reward anyway; a third group was promised nothing and got nothing.

The important result of the study was that the group that was promised the reward (and got it) was less interested in continuing the activity during a subsequent "free-choice" period compared to the no-reward control group. Lepper's explanation of these results was in line with "cognitive constructivist" theories of emotion and motivation that were popular in social psychology at the time. To put it briefly, the children who were promised the "Good Player" award attributed their drawing behavior to the promise of the reward; so when the reward was no longer offered, they were no longer interested in the activity.

This is what Lepper and Greene (1978) called "the hidden cost" of reward, and this argument has played a role ever since in debates over whether, for example, students should be paid for attending classes, completing their assignments on time, getting good grades, performing community service - -what have you.

The Lepper study is often cited as an argument against the use of rewards as instruments of social policy. This argument was repeated by Alfie Kohn in Punished by Rewards: The Trouble with Gold Stars, Incentive Plans, A's, Praise, and Other Bribes (1993). Nevertheless, a number of school districts and educational reform groups, as well as some private tutoring services, have proposed giving children money or other prizes as rewards for attending school and performing well. When, in 2007, the New York City Department of Education proposed rewarding students for school attendance and exam performance, in an attempt to improve educational outcomes for minority and other disadvantaged students, Barry Schwartz, a psychologist, cited Lepper's research in a New York Times Op-Ed piece arguing against the plan.

Nevertheless, such programs are increasingly popular. Of course, children around the world get their allowances contingent on cutting the grass or drying the dishes. But there's more: Schoolchildren are rewarded with cash and prizes for getting good grades, reading books, even just attending class and behaving properly; Homeowners are rewarded for recycling; Smokers are rewarded for cutting down or quitting; Employees are rewarded for skipping the French fries in the factory cafeteria; Adolescent girls are rewarded for not getting pregnant; Patients are rewarded for taking their medicine (examples from "The Age of Incentives: Paying Big Bucks for Puny Results" by Eric Felten, Wall Street Journal, 06/18/2010). Mainstream, market-based economists often favor proposals like these. But here are legitimate debates about the ethics of rewarding good behavior. And, as Lepper suggests, there are psychological reasons to think that they might
backfire.

But, and this is getting to be a theme with me, things are not quite this simple. Among the first to notice this was Judith Harackiewicz, who I'm proud to say was also my very first graduate student, at Harvard. Judy noticed what others had failed to, which was that the children who got the unexpected reward didn't show a decline in intrinsic motivation -- if anything, it increased. So, it's not reward, exactly, that undermines intrinsic motivation -- it's what the reward signifies.

Harackiewicz, along with her students and colleagues, subsequently embarked on an extensive program of research intended to disentangle the complicated effects of reward on intrinsic motivation. These experiments, like Dweck's, have moved fluidly from laboratory to field settings, using tasks that are truly intrinsically motivating for her subjects.

In the first place, Harackiewicz and her colleagues argue that we have to consider the **structure of reward**. Some rewards are **task-contingent**, in that they depend only on whether the person engages in some activity. This was the kind of reward used by Lepper et al. when they offered "Good Player awards just for drawing on paper. Other rewards are **performance-contingent**, in that they require the person to meet some specified standard of performance.

There is also the matter of **evaluative contingency**: Does the person expect to receive the reward from the outset, or is the person surprised by the reward after completing the task.

By their very nature, performance-contingent rewards provide **feedback** to the person about the quality of his performance. However, it's possible to provide feedback without any accompanying reward (coaches do this all the time during practice sessions).

And then there is the **delivery** of the reward. Rewards have symbolic cue value, because they represent the fact that you completed a task or did well. But they're also something tangible, like a trophy you can hold in your hand or an amount of cash you can spend.

Moreover, we have to consider the **type of reward**. Some rewards are **controlling**, in that they are incentives intended to get a person to engage in the task at all, or to perform at a particular standard, regardless of what they really want. Others are strictly **informational**, in that they communicate to the person (and others) How well he or she has done.

It turns out that these various aspects of reward make a big difference to the effects of rewarding behavior. This was clearly demonstrated in a study of college students who were brought into the laboratory to do something that was intrinsically motivating for them -- playing pinball; as a reward, they received movie passes for
achieving a meaningful but reasonable-sounding standard of performance (scoring above the 50th or 80th percentile); and rigging the machine to make sure that every one of the students met this standard. At the end of the experiment, all subjects were given an opportunity to continue playing the game during some free time, and the experimenters measured how long they continued playing.

(Let me just inject, as a point of personal privilege, that I keep in my head a "Faustian" list of experiments I would sell my soul to have done. It's a short list, and this one is on it. It's a real thing of beauty.)

In their first experiment, a promised reward undermined intrinsic motivation, compared to a standard control group that got performance feedback (i.e., they saw their score) but no evaluation and no reward. A third group that was surprised with the reward showed an enhancement of intrinsic motivation. This essentially replicates Lepper's original experiment, but with college students rather than nursery-school pupils. The promised reward could be perceived as controlling behavior, and probably was; and the subjects' anticipation of external evaluation probably induced performance anxiety. On the other hand, the unexpected reward was purely informational. The outcome suggests that controlling rewards undermine intrinsic motivation while purely informational rewards sustain, and may even enhance it.

In their second experiment, some subjects got the evaluation -- they were told that they had met the standard -- but received no reward. Again, this condition was intended to increase evaluation apprehension and performance anxiety, without the controlling element introduced by the offer of the movie tickets (the students got the movie tickets anyway, after they had completed their "free play" period). And again, this condition undermined intrinsic motivation, compared to the standard control group. But again, intrinsic motivation was enhanced for those students who were surprised with the purely informational unexpected reward (because there was no standard set or reward promised before, it was hoped that these subjects would not experience any evaluation apprehension or performance anxiety).

In a third experiment, again, some subjects got evaluative feedback but no reward offer; and, again, they showed diminished intrinsic motivation compared to the standard control group. Again, this illustrates the deleterious effects of evaluation apprehension and performance anxiety. But in this study the third group of subjects received information about normative performance -- they were told what the 80th percentile was -- but they got no promise of reward nor any hint of external evaluation (they did receive the reward, though, as a surprise). In this condition, they maintained or enhanced intrinsic motivation. So, if you can do it without seeming controlling, and without generating anxiety over performance evaluation, information about performance enhances intrinsic motivation, and giving a reward
doesn't compromise it.

There are some complications and subtleties that I don't have time to get into, but the bottom line on reward seems to be this: Rewards that are perceived as controlling behavior tend to undermine intrinsic motivation. But rewards that are informative about the person's level of competence, without being perceived as controlling, maintain and even enhance intrinsic motivation. The short story is that if we're going to offer rewards, the rewards ought to be offered for competence.

Based on this whole line of research, Harackiewicz and Sansone (1991, 2000) have explored a detailed psychological model that takes account of the factors that undermine, support, and even enhance intrinsic motivation.

- First they take account of **contextual and personality factors**, such as whether the reward is task-contingent or performance-contingent, the individual's initial level of interest in the task, and his or her orientation toward achievement in general.
- **Perceived purpose goals** establish the motivational context for the activity - what the person is trying to accomplish with the task -- to develop or demonstrate competence, to have fun, to relax, to socialize?
- While purpose goals reflect the reason(s) why the person is engaged in the activity, **perceived target goals** -- reflect exactly what it is that the person trying to accomplish -- for example, to score more than 20,000 points.
- Various **motivational processes** are also important: How much does the person value achieving (or displaying) competence in the task? Is he really involved in what he is doing, or is he just going through the motions? And does he perceive himself as competent -- never mind what the objective fact of the matter is, it's self-perception that matters most.

All of these factors combine and interact to determine the level of a person's intrinsic motivation to engage in some activity.

Motivation and reward are not too important for learning, as we've known since the animal studies of Tolman and Harlow, but they're obviously important for behavior. But extrinsic motives are not the sole determinants of behavior. Intrinsic motives are also important, and extrinsic motives do not always undermine intrinsic motives. The effects of reward on intrinsic motivation depend on what the reward is for, how the reward is perceived, and whether the person cares about the reward.

### Caring About Learning

Also relevant is whether the person cares about **learning**. We'd like to think that
students come to our classes intrinsically interested in what we have to teach them. That's often true -- though there are nasty things like prerequisites and distribution requirements that may be seen as purely instrumental (Hidi and Harackiewicz, 2000).

Suzanne Hidi and her colleagues have distinguished two quite different kinds of academic interest: **Enduring** interest is an individual, trait-like characteristic that is relatively stable over time. **Situational** interest is more specific to the student's immediate surroundings, and it's more transient. Hidi and Baird (1988) have argued, however, that situational interest can develop into enduring interest. After all, enduring interest in some subject matter has to come from somewhere, and it's not likely to be found in the genes. Once acquired, this enduring interest can then direct the student along relevant educational and career pathways.

The kind of interest a student has in a course is important, because it can affect the goals that the student adopts. For example, Harackiewicz and her colleagues identified three quite different groups of students taking an introductory psychology course. One group was just intrinsically interested in the material, and were taking the course for no other reason. Another group's interest was purely instrumental: they were taking the course to fulfill a distribution requirement. A third group had both sources of interest: they weren't intending to major in psychology, but they were interested enough in the course material that they would have taken it anyway, even if it hadn't been required. Students who were intrinsically interested in the course adopted mastery goals -- they really wanted to learn how the mind works. Students who were simply fulfilling a distribution requirement adopted performance goals -- they just want to get through the course, and don't want to hurt their GPA.

Hidi and Renninger (2006) have argued that interest can be stimulated, and maintained, by a number of determinants:

- **positive affect** aroused by the learning situation;
- the development of **content knowledge**;
- and the **perceived value** of what the student is learning.

But it's also true, unfortunately, that some students lose interest in what they're studying; and in other cases, interest just never catches fire. Longitudinal studies show that students' trajectory of interest slopes downhill as the academic term goes on -- and this is as true in college as it is in elementary and high school.

Here are some results from a study by Harackiewicz and her colleagues of college students taking an introductory psychology course. Almost every student comes to college intending to take introductory psychology, so interest starts out pretty high, but once the students find out that they're not going to find out how to get dates, manage their parents, and diagnose their roommates -- at least, not yet -- interest starts to decline. But in this study Hulleman, Harackiewicz, and their
colleagues (2010) tested a brief and remarkably simple intervention. One group of subjects was simply assigned to write a little essay about how the course material was relevant to their lives; a control group simply wrote a summary of the material learned so far. In this course, Harackiewicz was not able to introduce the "value intervention" until after the first midterm, by which time the students' interest trajectory had already started to slide downhill. Students who did well on the midterm maintained their level of interest thereafter, regardless of the intervention. But the students who did poorly, but who also got the "value intervention", also maintained their interest in the course. The students who did poorly on the midterm, and who were in the control group -- they continued to slide down.

For those of you who might want to try this for yourselves, the writing assignment was included as part of the syllabus, and students received some course credit for completing it. In fact, there were two relevance manipulations. In one, the students were asked to write a letter to a friend, or other significant person in their lives, about the relevance to their lives of some topic covered in the course. In the other, they were asked to write about the relevance to their lives of some course-relevant media report.

Note that this was a one-time intervention -- one relatively brief essay. The implication is that, had the investigators been able to intervene at the outset, the downward trajectory might have been prevented completely. But the more important implication is that, if you want to maintain student interest in a course, it's helpful for them to remind themselves why they're taking it.

Here's another example from Harackiewicz and her colleagues, who are particularly interested in getting high-school and college students interested in the so-called STEM disciplines of science, technology, engineering, and mathematics. In this study, Hulleman and Harackiewicz (2009) gave students in 9th and 10th-grade science classes a simple writing assignment. On several different occasions during the term, some students were asked to write short essays concerning the relevance of the course material to their own lives -- for example, their hoped-for future careers; a control group was asked merely to summarize the course material. Later on, the investigators gathered measures of interest in science and plans for taking science courses in the future, as well as course grades. The writing manipulation had little effect on those students who, for whatever reason, were already interested in science, and expected to do well. But for those students with low expectations of their own performance, the manipulation boosted both interest and grades. Call this the relevance effect.

The relevance effect works with students, but it can work through parents, as well. In a recent study Harackiewicz and her colleagues sent the parents of 10th- and 11th-grade high-school students a brochure, as well as a link to a website, which stressed the importance of math and science courses to both their children's daily life and their future

careers. They also included information about how parents could talk to their offspring about the STEM disciplines, and especially how to make connections between science and math and adolescents' lives. The manipulation had a substantial impact on students' registration for later, advanced, elective courses in STEM disciplines -- even for those students whose parents, themselves, had education beyond a bachelor's degree.

What Is To Be Done?

The nice thing about being a teacher is that learning comes naturally to our students. As Berkeley's Allison Gopnik is fond of pointing out, we appear to be \textit{wired to learn} -- right out of the womb, and maybe even in the womb, we're picking up information about our environment, calculating correlations and contingencies, in a naive but essentially scientific process of experimentation and statistical analysis, formulating, testing, and revising theories about how we and the world work. This attempt to predict and control, what the British psychologist F.C. Bartlett called \textit{effort after meaning}, is the essence of the learning process.

One radical inference from this view is that we erstwhile teachers should just \textit{get out of the way}. But that's not the right way to think about it. Learning comes naturally, but teachers can play a role in creating an optimal environment for learning to occur. As cognitive psychologists, we know a lot about how students learn, and how they forget and remember what they've learned. But as social psychologists we know that learning and memory occur in a personal and social context -- what I like to call the \textit{human ecology} of learning and memory. That includes the student's values, goals, and motives; and it also includes the interpersonal and institutional framework in which the individual student's learning activities take place. And, bringing us back to cognition, much depends on how these social factors are \textit{perceived}. The interests, values, goals, and motives that students bring to the learning environment are at least as important as the abilities and strategies that they bring to the task of learning.

So, in the spirit of distributed practice, let me summarize a set of teaching principles that promote effective learning. In this, I'm paraphrasing, sometimes directly quoting, the recommendations of a blue-ribbon panel convened by the National Center for Education Research in 2007 (Pashler et al., 2007; see also Bransford et al., 2000; Graesser, 2011)
1. **Space learning over time.** Shorter study sessions, interspersed with other activities, yield better long-term retention than the same amount of study all at once. Once a teacher has identified the key facts, terms, concepts, and skills to be learned, students to be exposed to each of them at least twice, separated by a period of several weeks, and arrange assignments and exams to promote distributed practice.

2. **Alternate between solved examples and problem sets.** Teachers can provide students with step-by-step solutions to sample problems, but they should also make sure that students have the opportunity to solve similar problems by themselves. At the very least, students should alternate between textbook and chalkboard examples that have already been worked out, and problems they must solve on their own, gradually decreasing the former and increasing the latter.

3. **Combine words and graphics.** Anything you can do to make study material richer will also make it more memorable. Pictures really are worth a thousand words -- even, in a literature course, "maps" of plotlines and the relationships among characters.

4. **Integrate the concrete with the abstract.** Illustrate abstract concepts with many and varied concrete examples.

5. **Testing promotes learning.** Not just midterms and finals, but also quizzes along the way. They ensure that students keep up with the material, but also aid spaced practice. Before introducing a new topic, prepare students with "pre-questions"; use quizzes to promote both retrieval practice and distributed practice.

6. **Help students allocate their time effectively.** Students are no better at managing their time than faculty are, and probably they're worse. A structured course, with deadlines and focused activities, will help a lot. Students are not particularly good at judging whether they've mastered a particular concept; at the very least, they should be taught to make these kinds of judgment after a delay has elapsed since learning. And teachers need to provide corrective feedback -- not just a quiz score or, much less, a checkmark in the upper right-hand corner of the first page of an assignment.

7. **Ask deep explanatory questions.** The examples I've given are usually in terms of facts, concepts, and skills, and they're important, but we're also after something deeper by way of understanding. Teachers should ask deep questions, encourage students to "think aloud" about the answers, and -- again -- provide feedback. Asking for explanations, as opposed to rote repetition or mere description, promotes elaborative processing, and thus improves long-term retention.
Link to a Practice Guide, "Organizing Instruction and Study to Improve Student Learning", published by the Institute of Education Sciences, from which these points were drawn.

Link to a list of 25 principles that aid student learning, assembled by Prof. Art Graesser of the University of Memphis, a leading investigator in this area.

Selected References


This page last modified 04/25/2013 9:11 AM.