**Why Flunking Exams Is Actually a Good Thing**

By [BENEDICT CAREY](http://www.nytimes.com/by/benedict-carey), *New York Times Magazine: The Education Issue*, SEPT. 4, 2014

Credit Illustration by Mark Pernice

Imagine that on Day 1 of a difficult course, before you studied a single thing, you got hold of the final exam. The motherlode itself, full text, right there in your email inbox — attached mistakenly by the teacher, perhaps, or poached by a campus hacker. No answer key, no notes or guidelines. Just the questions.

Would that help you study more effectively? Of course it would. You would read the questions carefully. You would know exactly what to focus on in your notes. Your ears would perk up anytime the teacher mentioned something relevant to a specific question. You would search the textbook for its discussion of each question. If you were thorough, you would have memorized the answer to every item before the course ended. On the day of that final, you would be the first to finish, sauntering out with an A+ in your pocket. And you would be cheating.

But what if, instead, you took a test on Day 1 that was just as comprehensive as the final but *not* a replica? You would bomb the thing, for sure. You might not understand a single question. And yet as disorienting as that experience might feel, it would alter how you subsequently tuned into the course itself — and could sharply improve your overall performance.

This is the idea behind pretesting, one of the most exciting developments in learning-­science. Across a variety of experiments, psychologists have found that, in some circumstances, wrong answers on a pretest aren’t merely useless guesses. Rather, the attempts themselves change how we think about and store the information contained in the questions. On some kinds of tests, particularly multiple-choice, we benefit from answering incorrectly by, in effect, priming our brain for what’s coming later.

That is: The (bombed) pretest drives home the information in a way that studying as usual does not. We fail, but we fail forward.

The excitement around prefinals is rooted in the fact that the tests appear to improve subsequent performance in topics that are not already familiar, whether geography, sociology or psychology. At least they do so in experiments in controlled laboratory conditions. A just-completed study — the first of its kind, carried out by the U.C.L.A. psychologist Elizabeth Ligon Bjork — found that in a live classroom of Bjork’s own students, pretesting raised performance on final-exam questions by an average of 10 percent compared with a control group.

The basic insight is as powerful as it is surprising: Testing might be the key to studying, rather than the other way around. As it turns out, a test is not only a measurement tool. It’s a way of enriching and altering memory.

**Many of us** dread tests because we’ve been wounded by a few over the years, and sometimes severely. Almost everyone has had at least one lost-in-space experience, opening an exam to find a long list of questions that seem to hail from another course altogether. Vision narrows, the mind seizes; all feeling drains from the extremities. We would crawl into a hole if we weren’t already in one.

Yet another species of exam collapse is far more common. These are the cases in which we open the test and see familiar questions on material we’ve studied, perhaps even stuff we’ve highlighted with yellow marker: names, ideas, formulas we could recite easily only yesterday. And still we lay an egg, scoring average or worse.

Why does this happen? Psychologists have studied learning long enough to have an answer, and typically it’s not a lack of effort (or of some elusive test-taking gene). The problem is that we have misjudged the depth of what we know. We are duped by a misperception of “fluency,” believing that because facts or formulas or arguments are easy to remember *right now,* they will remain that way tomorrow or the next day. This fluency illusion is so strong that, once we feel we have some topic or assignment down, we assume that further study won’t strengthen our memory of the material. We move on, forgetting that we forget.

Often our study “aids” simply create fluency illusions — including, yes, highlighting — as do chapter outlines provided by a teacher or a textbook. Such fluency misperceptions are automatic; they form subconsciously and render us extremely poor judges of what we need to restudy or practice again. “We know that if you study something twice, in spaced sessions, it’s harder to process the material the second time, and so people think it’s counterproductive,” Nate Kornell, a psychologist at Williams College, said. “But the opposite is true: You learn more, even though it feels harder. Fluency is playing a trick on judgment.”

The best way to overcome this illusion is testing, which also happens to be an effective study technique in its own right. This is not exactly a recent discovery; people have understood it since the dawn of formal education, probably longer. In 1620, the philosopher Francis Bacon wrote, “If you read a piece of text through twenty times, you will not learn it by heart so easily as if you read it ten times while attempting to recite it from time to time and consulting the text when your memory fails.”

Scientific confirmation of this principle began in 1916, when Arthur Gates, a psychologist at Columbia University, created an ingenious study to further Bacon’s insight. If someone is trying to learn a piece of text from memory, Gates wondered, what would be the ideal ratio of study to recitation (without looking)? To interrogate this question, he had more than 100 schoolchildren try to memorize text from Who’s Who entries. He broke them into groups and gave each child nine minutes to prepare, along with specific instructions on how to use that time. One group spent 1 minute 48 seconds memorizing and the remaining time rehearsing (reciting); another split its time roughly in half, equal parts memorizing and rehearsing; a third studied for a third and recited for two-thirds; and so on.

After a sufficient break, Gates sat through sputtered details of the lives of great Americans and found his ratio. “In general,” he concluded, “best results are obtained by introducing recitation after devoting about 40 percent of the time to reading. Introducing recitation too early or too late leads to poorer results.” The quickest way to master that Shakespearean sonnet, in other words, is to spend the first third of your time memorizing it and the remaining two-thirds of the time trying to recite it from memory.

In the 1930s, a doctoral student at the State University of Iowa, Herbert F. Spitzer, recognized the broader implications of this insight. Gates’s emphasis on recitation was, Spitzer realized, not merely a study tip for memorization; it was nothing less than a form of self-examination. It was testing as study, and Spitzer wanted to extend the finding, asking a question that would apply more broadly in education: If testing is so helpful, when is the best time to do it?

He mounted an enormous experiment, enlisting more than 3,500 sixth graders at 91 elementary schools in nine Iowa cities. He had them study an age-appropriate article of roughly 600 words in length, similar to what they might analyze for homework. Spitzer divided the students into groups and had each take tests on the passages over the next two months, according to different schedules. For instance, Group 1 received one quiz immediately after studying, then another a day later and a third three weeks later. Group 6, by contrast, didn’t take one until three weeks after reading the passage. Again, the time the students had to study was identical. So were the quizzes. Yet the groups’ scores varied widely, and a clear pattern emerged.

The groups that took pop quizzes soon after reading the passage — once or twice within the first week — did the best on a final exam given at the end of two months, marking about 50 percent of the questions correct. (Remember, they had studied their peanut or bamboo article only once.) By contrast, the groups who took their first pop quiz two weeks or more after studying scored much lower, below 30 percent on the final. Spitzer’s study showed that not only is testing a powerful study technique, but it’s also one that should be deployed sooner rather than later. “Achievement tests or examinations are learning devices and should not be considered only as tools for measuring achievement of pupils,” he concluded.

The testing effect, as it’s known, is now well established, and it opens a window on the alchemy of memory itself. “Retrieving a fact is not like opening a computer file,” says Henry Roediger III, a psychologist at Washington University in St. Louis, who, with Jeffrey Karpicke, now at Purdue University, has established the effect’s lasting power. “It alters what we remember *and* changes how we subsequently organize that knowledge in our brain.”

**If tests are** most effective when given sooner rather than later, then why not go the distance? Why not give the final on the first day, as well as on the last? This is the radical question that Bjork, the U.C.L.A. psychologist, has set out to investigate.

She did not actually give a comprehensive prefinal on the first day of class, in order to avoid overwhelming her students. She also decided to start with fairly basic material, conducting the study on her Psychology 100B class at U.C.L.A., which covers research methods.

She and Nicholas Soderstrom, a postdoc, gave the entire class of more than 300 students a short pretest, all multiple-choice questions, immediately before the start of some lectures but not others. “We wanted to see whether students would better remember and understand material from lectures preceded by a pretest than from lectures not preceded by a pretest,” Soderstrom said.

To answer that, Bjork and Soderstrom did something clever on a cumulative final exam, which was given at the end of the course. Namely, they included on it questions that were related to the pretest ones as well as questions that were not. “If pretesting helps, then students should do better on related questions during a later exam than on questions about material we covered in the lectures but was not pretested,” Bjork said. She and Soderstrom would compare students’ scores on pretest-­related questions with their scores on nonpretested ones, to see if there was any difference.

For example, here’s a question from one of the pretests:

**Which of the following is true of scientific explanations?**

**a.** They are less likely to be verified by empirical observation than other types of explanations.

**b.** They are accepted because they come from a trusted source or authority figure.

**c.** They are accepted only provisionally.

**d.** In the face of evidence that is inconsistent with a scientific explanation, the evidence will be questioned.

**e.** All of the above are true about scientific explanations.

And here’s a related question, from the cumulative test given after the lectures:

**Which of the following is true of explanations based on belief?**

**a.** They are more likely to be verified by empirical observation than other types of explanations.

**b.** They are accepted because they come from a trusted source or authority figure.

**c.** They are assumed to be true absolutely.

**d.** In the face of evidence that is inconsistent with an explanation based on belief, the belief will be questioned.

**e.** b and c above.

The students tanked all three pretests, performing no better than if they had guessed at random. Bjork and Soderstrom had expected as much. But the class received prompt feedback, attending the relevant lecture shortly after they took each of the three pretests. Those lectures in effect supplied them with correct answers to questions that had just been posed on the pretest. In previous experiments, such immediacy seemed to be a critical component: Pretests led to the most improvement when students received the correct answers reasonably soon after their guessing.

In order to gauge the effect of the testing, Bjork and Soderstrom gave a cumulative exam at the end of the 10-week course. It was the same format as the others: multiple-choice questions, each with five possible answers. The result? Bjork’s Psych 100B class scored about 10 percent higher on the related questions than on the unrelated ones. It’s far from a magic memory pill — but 10 percent, as we all know, can often translate to a letter grade. “On the basis of this significant difference,” Bjork said, “giving students a pretest on topics to be covered in a lecture improves their ability to answer related questions about those topics on a later final exam.” Even when students bomb, she said, pretests provide them an opportunity to see what vocabulary will be used in the coming lectures, what kinds of questions will be posed and which distinctions between concepts will be crucial.

**Bjork’s experiment** suggests that pretesting serves to prime the brain, predisposing it to absorb new information. Scientists have several theories as to how this happens. One is fairly obvious: Students get a glimpse from a pretest of the teacher’s hand, of what they’ll be up against. That’s in the interest of not just students but of teachers, too. You can teach facts and concepts all you want, but what’s most important in the end is how students *think* about that material: How they incorporate all those definitions into a working narrative about a topic that gives them confidence in judging what’s important and what’s less so. These are not easy things to communicate, even for the best teachers. You can’t download such critical thinking quickly, hard as you might try. But you can easily give a test with questions that themselves force that kind of hierarchical thinking. “Taking a practice test and getting wrong answers seems to improve subsequent study, because the test adjusts our thinking in some way to the kind of material we need to know,” Bjork said.

A second possibility has to do with the concept of fluency. Wrong guesses expose our fluency illusions, our false impression that we “know” the capital of Eritrea because we just saw it or once studied it. A test, if multiple-choice, forces us to select the correct answer from a number of possibilities that also look plausible. “Let’s say you’re pretty sure that Australia’s capital is Canberra,” Robert A. Bjork, Elizabeth Ligon Bjork’s husband and a leading learning scientist, said. “O.K., that seems easy enough. But when the exam question appears, you see all sorts of other possibilities — Sydney, Melbourne, Adelaide — and suddenly you’re not so sure. If you’re studying just the correct answer, you don’t appreciate all the other possible answers that could come to mind or appear on the test.” Pretesting operates as a sort of fluency vaccine.

Biologically, too, there may be something deeper at work. To review, memory builds on itself in ways we don’t usually notice. Retrieval — i.e. remembering — is a different mental act than straight studying; the brain is digging out a fact, together with a network of associations, which alters and enriches how that network is subsequently re-stored. But guessing is distinct from both study and retrieval. It too will reshape our mental networks by embedding unfamiliar concepts (the lend-lease program, the confirmation bias, the superego) into questions we at least partly comprehend (“Name one psychological phenomenon that skews our evaluation of evidence”). Even if the question is not entirely clear and its solution unknown, a guess will in itself begin to link the questions to possible answers. And those networks light up like Christmas lights when we hear the concepts again.

And here is where pretesting shows its likely limitations: A prefinal for an intro class in Arabic or Chinese could be a wash, because the notations and characters are entirely alien. There’s no scaffolding of familiar language to work with — no existing network in which to situate the new symbols — before we make a guess. We are truly lost, with no recognizable landmark. The research thus far suggests that prefinals will be much more useful in humanities courses and social-science disciplines in which unfamiliar concepts are at least embedded in language we can parse.

The word “testing” is still loaded, of course, in ways that have nothing to do with learning science. Educators and experts have debated the value of standardized testing for decades, and reforms like the No Child Left Behind law, which increased the use of such exams, have only inflamed the argument. Many teachers complain that a focus on testing limits their ability to fully explore subjects with their students. Others attack tests as woefully incomplete measures of learning, blind to all varieties of creative thinking.

But the emerging study of pretesting flips that logic on its head. “Teaching to the test” becomes “learning to understand the pretest,” whichever one the teacher chooses to devise. The test, that is, becomes an introduction to what students should learn, rather than a final judgment on what they did not.

**Correction: October 26, 2014**

An article on Sept. 7 about pretesting as a learning technique misstated the given name of the author of a 1939 psychological paper on the retention of knowledge. He was Herbert F. Spitzer, not Herman.

[Benedict Carey](mailto:bencarey@nytimes.com) is a science reporter for The Times and the author of “How We Learn: The Surprising Truth About When, Where and Why It Happens,” to be published by Random House this month, from which this article is adapted.

Editor: [Ilena Silverman](mailto:ilenas@nytimes.com)

A version of this article appears in print on September 7, 2014, on page MM38 of the Sunday Magazine with the headline: Test Drive. [Today's Paper](http://www.nytimes.com/pages/todayspaper/index.html" \t "_blank)|[Subscribe](http://www.nytimes.com/subscriptions/Multiproduct/lp839RF.html?campaignId=48JQY" \t "_blank)