

Excerpt from Benedict Carey's
book "How We Learn,"
which I highly recommend.

Chapter Five

The Hidden Value of Ignorance

The Many Dimensions of Testing

At some point in our lives, we all meet the Student Who Tests Well Without Trying. "I have no idea what happened," says she, holding up her 99 percent score. "I hardly even studied." It's a type you can never entirely escape, even in adulthood, as parents of school-age children quickly discover. "I don't know what it is, but Daniel just scores off the charts on these standardized tests," says Mom—dumbfounded!—at school pickup. "He certainly doesn't get it from me." No matter how much we prepare, no matter how early we rise, there's always someone who does better with less, who magically comes alive at game time.

I'm not here to explain that kid. I don't know of any study that looks at test taking as a discrete, stand-alone skill, or any evidence that it is an inborn gift, like perfect pitch. I don't need research to tell me that this type exists; I've seen it too often with my own eyes. I'm also old enough to know that being jealous isn't any way to close the gap between us and them. Neither is working harder. (Trust me, I've already tried that.)

No, the only way to develop any real test taking mojo is to understand more deeply what, exactly, testing *is*. The truth is not so self-evident, and it has more dimensions than you might guess.

The first thing to say about testing is this: Disasters happen. To everyone. Who hasn't opened a test booklet and encountered a list of questions that seem related to a different course altogether? I have a favorite story about this, a story I always go back to in the wake of any collapse. The teenage Winston Churchill spent weeks preparing for the entrance exam into Harrow, the prestigious English boys school. He wanted badly to get in. On the big day, in March of 1888, he opened the exam and found, instead of history and geography, an unexpected emphasis on Latin and Greek. His mind went blank, he wrote later, and he was unable to answer a single question. "I wrote my name at the top of the page. I wrote down the number of the question, '1.' After much reflection I put a bracket round it, thus, '(1)'. But thereafter I could not think of anything connected with it that was either relevant or true. Incidentally there arrived from nowhere in particular a blot and several smudges. I gazed for two whole hours at this sad spectacle; and then merciful ushers collected up my piece of foolscap and carried it up to the Headmaster's table."

And that's *Winston Churchill*.

The next thing to say is less obvious, though it's rooted in a far more common type of blown test. We open the booklet and see familiar questions on material we've studied, stuff we've highlighted with yellow marker: names, ideas, formulas we could recite with

ease only yesterday. No trick questions, no pink elephants, and still we lay an egg. Why? How? I did so myself on one of the worst possible days: a trigonometry final I needed to ace to get into an Advanced Placement course, junior year. I spent weeks preparing. Walking into the exam that day, I remember feeling pretty good. When the booklets were handed out, I scanned the questions and took an easy breath. The test had a few of the concepts I'd studied, as well as familiar kinds of questions, which I'd practiced dozens of times.

I can do this, I thought.

Yet I scored somewhere in the low 50s, in the very navel of average. (These days, a score like that would prompt many parents to call a psychiatrist.) Who did I blame? Myself. I knew the material but didn't hear the music. I was a "bad test taker." I was kicking myself—but for all the wrong reasons.

The problem wasn't that I hadn't worked hard enough, or that I lacked the test taking "gene." No, my mistake was misjudging the depth of what I knew. I was duped by what psychologists call fluency, the belief that because facts or formulas or arguments are easy to remember *right now*, they'll remain that way tomorrow or the next day. The fluency illusion is so strong that, once we feel we've nailed some topic or assignment, we assume that further study won't help. We forget that we forget. Any number of study "aids" can create fluency illusions, including (yes) highlighting, making a study guide, and even chapter outlines provided by a teacher or a textbook. Fluency misperceptions are automatic. They form subconsciously and make us 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," as Nate Kornell, a psychologist at Williams College, told me. "But the opposite is true: You learn more, even though it feels harder. Fluency is playing a trick on judgment."

So it is that we end up attributing our poor test results to "test anxiety" or—too often—stupidity.

Let's recall the Bjorks' "desirable difficulty" principle: The harder your brain has to work to dig out a memory, the greater the increase in learning (retrieval and storage strength). Fluency, then, is the flipside of that equation. The *easier* it is to call a fact to mind, the smaller the increase in learning. Repeating facts right after you've studied them gives you nothing, no added memory benefit.

The fluency illusion is the primary culprit in below-average test performances. Not anxiety. Not stupidity. Not unfairness or bad luck.

Fluency.

The best way to overcome this illusion and improve our testing skills is, conveniently, an effective study technique in its own right. The technique is not exactly a recent invention; people have been employing it since the dawn of formal education, probably longer. Here's the philosopher Francis Bacon, spelling it out in 1620: "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." And here's the irrepressible William James, in 1890, musing about

the same concept: “A curious peculiarity of our memory is that things are impressed better by active than by passive repetition. I mean that in learning—by heart, for example—when we almost know the piece, it pays better to wait and recollect by an effort from within, than to look at the book again. If we recover the words in the former way, we shall probably know them the next time; if in the latter way, we shall very likely need the book once more.”

The technique is testing itself. Yes, I am aware of how circular this logic appears: better testing through testing. Don’t be fooled. There’s more to self-examination than you know. A test is not only a measurement tool, it alters what we remember and *changes* how we subsequently organize that knowledge in our minds. And it does so in ways that greatly improve later performance.

• • •

One of the first authoritative social registries in the New World was *Who’s Who in America*, and the premiere volume, published in 1899, consisted of more than 8,500 entries—short bios of politicians, business leaders, clergymen, railroad lawyers, and sundry “distinguished Americans.” The bios were detailed, compact, and historically rich. It takes all of thirty seconds, for example, to learn that Alexander Graham Bell received his patent for the telephone in 1876, just days after his twenty-ninth birthday, when he was a professor of vocal physiology at Boston University. And that his father, Alexander Melville Bell (the next entry), was an inventor, too, an expert in elocution who developed Visible Speech, a set of symbols used to help deaf people learn to speak. And that *his* father—Alexander Bell, no middle name, of Edinburgh—pioneered the treatment of speech impediments. Who knew? The two younger Bells, though both were born in Edinburgh, eventually settled in Washington, D.C. The father lived at 1525 35th Street, and the son at 1331 Connecticut Avenue. That’s right, the addresses are here, too. (Henry James: Rye, Isle of Wight.)

In 1917, a young psychologist at Columbia University had an idea: He would use these condensed life entries to help answer a question. Arthur Gates was interested in, among other things, how the act of recitation interacts with memory. For centuries, students who received a classical education spent untold hours learning to recite from memory epic poems, historic monologues, and passages from scripture—a skill that’s virtually lost today. Gates wanted to know whether there was an ideal ratio between reading (memorizing) and reciting (rehearsal). If you want to learn Psalm 23 (*The Lord is my shepherd, I shall not want ...*) by heart—in, say, a half hour—how many of those minutes should you spend studying the verse on the page, and how many should you spend trying to recite from memory? What ratio anchors that material in memory most firmly? That would have been a crucial percentage to have, especially back when recitation was so central to education. The truth is, it’s just as handy today, not only for actors working to memorize Henry V’s St. Crispin’s Day speech but for anyone preparing a presentation, learning a song, or studying poetry.

To find out if such a ratio existed, Gates enlisted five classes from a local school,

ranging from third to eighth grade, for an experiment. He assigned each student a number of *Who's Who* entries to memorize and recite (the older students got five entries, the youngest ones three). He gave them each nine minutes to study along with specific instructions on how to use that time: One group would spend a minute and forty-eight seconds memorizing, and seven minutes, twelve seconds rehearsing (reciting); another would split its time in half, equal parts memorizing and rehearsing; a third, eight minutes of its time memorizing, and only a minute rehearsing. And so on.

Three hours later, it was showtime. Gates asked each student to recite what he or she could remember of their assigned entries:

“Edgar Mayhew Bacon, author ... born, uh, June 5, 1855, Nassau, the Bahamas, and uh, went to private schools in Tarrytown, N.Y.; worked in a bookstore in Albany, and then I think became an artist ... and then wrote, ‘The New Jamaica’? ... and ‘Sleepy Hollow’ maybe?”

One, after another, after another. Edith Wharton. Samuel Clemens. Jane Addams. The brothers James. More than a hundred students, reciting.

And in the end, Gates had his ratio.

“In general,” he concluded, “the 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,” Gates wrote. In the older grades, the percentage was even smaller, closer to a third. “The superiority of optimal reading and retention over reading alone is about 30 percent.”

The quickest way to download that St. Crispin’s Day speech, in other words, is to spend the first third of your time memorizing it, and the remaining two thirds reciting from memory.

Was this a landmark finding? Well, yes, actually. In hindsight, it was the first rigorous demonstration of a learning technique that scientists now consider one of the most powerful of all. Yet at the time no one saw it. This was one study, in one group of schoolchildren. Gates didn’t speculate on the broader implications of his results, either, at least not in the paper he published in the *Archives of Psychology*, “Recitation as a Factor in Memorizing,” and the study generated little scientific discussion or follow-up.

The reasons for this, I think, are plain enough. Through the first half of the twentieth century, psychology was relatively young and growing by fits and starts, whipsawed by its famous theorists. Freud’s ideas still cast a long shadow and attracted hundreds of research projects. Ivan Pavlov’s experiments helped launch decades of research on conditioned learning—stimulus-response experiments, many of them in animals. Research into education was in an exploratory phase, with psychologists looking into reading, into learning disabilities, phonics, even the effect of students’ emotional life on grades. And it’s important to say that psychology—like any science—proceeds in part by retrospective clue gathering. A scientist has an idea, a theory, or a goal, and looks backward to see if there’s work to build on, if there’s anyone who’s had the same idea or reported results that are supportive of it. Science may be built on the shoulders of

giants, but for a working researcher it's often necessary to ransack the literature to find out who those giants are. Creating a rationale for a research project can be an exercise in historical data mining—in finding shoulders to build on.

Gates's contribution is visible only in retrospect, but it was inevitable that its significance would be noticed. Improving education was, then as now, a subject of intense interest. And so, in the late 1930s, more than twenty years later, another researcher found in Gates's study a rationale for his own. Herbert F. Spitzer was a doctoral student at the State University of Iowa, who in 1938 was trawling for a dissertation project. He wasn't interested in recitation per se, and he didn't belong to the small club of academic psychologists who were focused on studying the intricacies of memory. He was intent on improving teaching methods, and one of the biggest questions hanging over teachers, from the very beginning of the profession, was *when* testing is most effective. Is it best to give one big exam at the end of a course? Or do periodic tests given earlier in the term make more sense?

We can only guess at Spitzer's thinking, because he did not spell it out in his writings. We know he'd read Gates's study, because he cites it in his own. We know, too, that he saw Gates's study for what it was. In particular, he recognized Gates's recitation as a form of self-examination. Studying a prose passage for five or ten minutes, then turning the page over to recite what you can without looking, isn't only practice. It's a test, and Gates had shown that that self-exam had a profound effect on final performance.

That is to say: Testing *is* studying, of a different and powerful kind.

Spitzer understood that, and then asked the next big question. If taking a test—whether recitation, rehearsal, self-exam, pop quiz, or sit-down exam—improves learning, then when is the best time to take it?

To try to find out, he mounted an enormous experiment, enlisting sixth graders at ninety-one different elementary schools in nine Iowa cities—3,605 students in all. He had them study an age-appropriate six-hundred-word article, similar to what they might get for homework. Some were assigned an article on peanuts, and others one on bamboo. They studied the passage once. Spitzer then divided the students into eight groups and had each group take several tests on the passages over the next two months. The tests for each group were all the same, multiple-choice, twenty-five questions, each with five possible answers. For example, for those who studied bamboo:

What usually happens to a bamboo plant after the flowering period?

- a. It dies
- b. It begins a new growth
- c. It sends up new plants from the roots
- d. It begins to branch out
- e. It begins to grow a rough bark

In essence, Spitzer conducted what was, and probably still is, the largest pop quiz experiment in history. The students had no idea that the quizzes were coming, or when. And each group got hit with quizzes at different times. Group 1 got one right after studying, then another a day later, and a third three weeks later. Group 6 didn't take their first quiz until three weeks after reading the passage. Again, the time the students had to study was identical. So were the questions on the quizzes.

Yet the groups' scores varied widely, and a 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, getting about 50 percent of the questions correct. (Remember, they'd 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 showed not only that testing is a powerful study technique, he showed it's one that should be deployed sooner rather than later.

“Immediate recall in the form of a test is an effective method of aiding the retention of learning and should, therefore, be employed more frequently,” he concluded. “Achievement tests or examinations are learning devices and should not be considered only as tools for measuring achievement of pupils.”

For lab researchers focused on improving retention, this finding should have rung a bell, and loudly. Recall, for a moment, Ballard's “reminiscence” from [chapter 2](#). The schoolchildren in his “Wreck of the Hesperus” experiment studied the poem only once but continued to improve on subsequent tests given days later, remembering more and more of the poem as time passed. Those intervals between studying (memorizing) the poem and taking the tests—a day later, two days, a week—are exactly the ones that Spitzer found most helpful for retention. Between them, Gates and Spitzer had demonstrated that Ballard's young students improved not by some miracle but because each test was an additional study session. Even then, after Spitzer published his findings in *The Journal of Educational Psychology*, the bell didn't sound.

“We can only speculate as to why,” wrote Henry Roediger III and Jeffrey Karpicke, also then at Washington University, in a landmark 2006 review of the testing effect, as they called it. One possible reason, they argued, is that psychologists were still primarily focused on the dynamics of forgetting: “For the purpose of measuring forgetting, repeated testing was deemed a confound, to be avoided.” It “contaminated” forgetting, in the words of one of Spitzer's contemporaries.

Indeed it did, and does. And, as it happens, that contamination induces improvements in thinking and performance that no one predicted at the time. More than thirty years passed before someone picked up the ball again, finally seeing the possibilities of what Gates and Spitzer had found.

That piece of foolscap Winston Churchill turned in, with the smudges and blots? It was far from a failure, scientists now know—even if he scored a flat zero.

• • •

Let's take a breather from this academic parsing of ideas and do a simple experiment, shall we? Something light, something that gets this point across without feeling like homework. I've chosen two short passages from one author for your reading pleasure—and pleasure it should be, because they're from, in my estimation, one of the most savage humorists who ever strode the earth, however unsteadily. Brian O'Nolan, late of Dublin, was a longtime civil servant, crank, and pub-crawler who between 1930 and 1960 wrote novels, plays, and a much beloved satirical column for *The Irish Times*. Now, your assignment: Read the two selections below, four or five times. Spend five minutes on each, then put them aside and carry on with your chores and shirking of same. Both come from a chapter called "Bores" in O'Nolan's book *The Best of Myles*:

Passage 1: The Man Who Can Pack

This monster watches you try to stuff the contents of two wardrobes into an attaché case. You succeed, of course, but have forgotten to put in your golf clubs. You curse grimly but your "friend" is delighted. He knew this would happen. He approaches, offers consolation and advises you to go downstairs and take things easy while he "puts things right." Some days later, when you unpack your things in Glengariff, you find that he has not only got your golf clubs in but has included your bedroom carpet, the kit of the Gas Company man who has been working in your room, two ornamental vases and a card-table. Everything in view, in fact, except your razor. You have to wire 7 pounds to Cork to get a new leather bag (made of cardboard) to get all this junk home.

Passage 2: The Man Who Soles His Own Shoes

Quite innocently you complain about the quality of present-day footwear. You wryly exhibit a broken sole. "Must take them in tomorrow," you say vaguely. The monster is flabbergasted at this passive attitude, has already forced you into an armchair, pulled your shoes off and vanished with them into the scullery. He is back in an incredibly short space of time and restored your property to you announcing that the shoes are now "as good as new." You notice his own for the first time and instantly understand why his feet are deformed. You hobble home, apparently on stilts. Nailed to each shoe is an inch-thick slab of "leather" made from Shellac, saw-dust and cement.

Got all that? It's not *The Faerie Queene*, but it'll suffice for our purposes. Later in the day—an hour from now, if you're going with the program—restudy Passage 1. Sit down for five minutes and reread it a few more times, as if preparing to recite it from memory (which you are). When the five minutes are up, take a break, have a snack, and come back to Passage 2. This time, instead of restudying, test yourself on it. Without looking, write down as much of it as you can remember. If it's ten words, great. Three sentences? Even better. Then put it away without looking at it again.

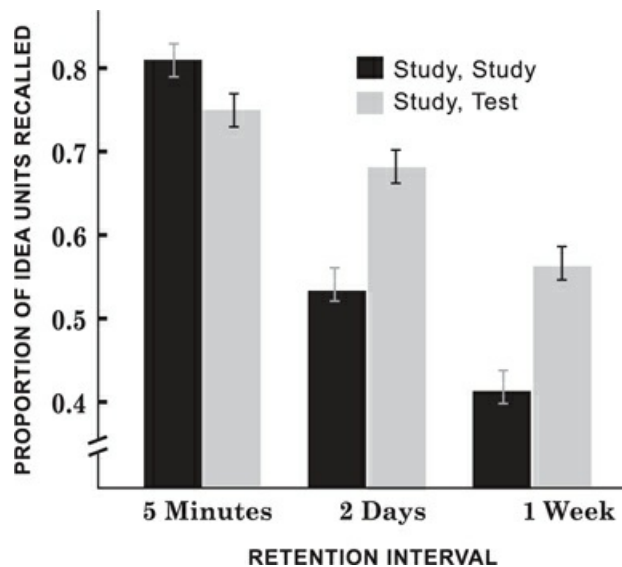
The next day, test yourself on both passages. Give yourself, say, five minutes on each to recall as much as you can.

So: Which was better?

Eyeball the results, counting the words and phrases you remembered. Without being there to look over your shoulder and grade your work, I'm going to hazard a guess that you did markedly better on the second passage.

That is essentially the experimental protocol that a pair of psychologists—Karpicke, now at Purdue, and Roediger—have used in a series of studies over the past decade or so. They've used it repeatedly, with students of all ages, and across a broad spectrum of material—prose passages, word pairs, scientific subjects, medical topics. We'll review one of their experiments, briefly, just to be clear about the impact of self-examination. In a 2006 study, Karpicke and Roediger recruited 120 undergraduates and had them study two science-related passages, one on the sun and the other on sea otters. They studied one of the two passages twice, in separate seven-minute sessions. They studied the other one once, for seven minutes, and in the next seven-minute session were instructed to write down as much of the passage as they could recall without looking. (That was the “test,” like we just did above with the O’Nolan passages.) Each student, then, had studied one passage two times—either the sea otters, or the sun—and the other just once, followed by a free recall test on it.

Karpicke and Roediger split the students into three groups, one of which took a test five minutes after the study sessions, one that got a test two days later, and one that tested a week later. The results are easily read off the following graph:



There are two key things to take away from this experiment. First, Karpicke and Roediger kept preparation time equal; the students got the same amount of time to try to learn both passages. Second, the “testing” prep *buried* the “study” prep when it really mattered, on the one-week test. In short, testing does not = studying, after all. In fact, testing > studying, and by a country mile, on delayed tests.

“Did we find something no one had ever found before? No, not really,” Roediger told me. Other psychologists, most notably Chizuko Izawa, had shown similar effects in the 1960s and ’70s at Stanford University. “People had noticed testing effects and gotten excited about them. But we did it with different material than before—the prose passages, in this case—and I think that’s what caught people’s attention. We showed that this could be applied to real classrooms, and showed how strong it could be. That’s when the research started to take off.”

Roediger, who’s contributed an enormous body of work to learning science, both in experiments and theory, also happens to be one of the field’s working historians. In a review paper published in 2006, he and Karpicke analyzed a century’s worth of experiments, on all types of retention strategies (like spacing, repeated study, and context), and showed that the testing effect has been there all along, a strong, consistent “contaminant,” slowing down forgetting. To measure any type of learning, after all, you have to administer a test. Yet if you’re using the test only for measurement, like some physical education push-up contest, you fail to see it as an added workout—itsself making contestants’ memory muscles stronger.

The word “testing” is loaded, in ways that have nothing to do with learning science. Educators and experts have debated the value of standardized testing for decades, and reforms instituted by President George W. Bush in 2001—increasing the use of such exams—only inflamed the argument. Many teachers complain of having to “teach to the test,” limiting their ability to fully explore subjects with their students. Others attack such tests as incomplete measures of learning, blind to all varieties of creative thinking. This debate, though unrelated to work like Karpicke and Roediger’s, has effectively prevented their findings and those of others from being applied in classrooms as part of standard curricula. “When teachers hear the word ‘testing,’ because of all the negative connotations, all this baggage, they say, ‘We don’t need more tests, we need *less*,’ ” Robert Bjork, the UCLA psychologist, told me.

In part to soften this resistance, researchers have begun to call testing “retrieval practice.” That phrase is a good one for theoretical reasons, too. If self-examination is more effective than straight studying (once we’re familiar with the material), there must be reasons for it. One follows directly from the Bjorks’ desirable difficulty principle. When the brain is retrieving studied text, names, formulas, skills, or anything else, it’s doing something different, and *harder*, than when it sees the information again, or restudies. That extra effort deepens the resulting storage and retrieval strength. We know the facts or skills better because we retrieved them ourselves, we didn’t merely review them.

Roediger goes further still. When we successfully retrieve a fact, he argues, we then re-store it in memory in a different way than we did before. Not only has storage level spiked; the memory itself has new and different connections. It’s now linked to other related facts that we’ve also retrieved. The network of cells holding the memory has itself been altered. Using our memory changes our memory in ways we don’t anticipate.

And that’s where the research into testing takes an odd turn indeed.

...

What if you somehow got hold of the final exam for a course on Day 1, before you'd even studied a thing? Imagine it just appeared in your inbox, sent mistakenly by the teacher. Would having that test matter? Would it help you prepare for taking the final at the end of the course?

Of course it would. You'd read the questions carefully. You'd know what to pay attention to and what to study in your notes. Your ears would perk up anytime the teacher mentioned something relevant to a specific question. If you were thorough, you'd have memorized the correct answer to every item before the course ended. On the day of that final, you'd be the first to finish, sauntering out with an A+ in your pocket.

And you'd be cheating.

But what if, instead, you took a test on Day 1 that was comprehensive but *not* a replica of the final exam? You'd bomb the thing, to be sure. You might not be able to understand a single question. And yet that experience, given what we've just learned about testing, might alter how you subsequently tune into the course itself during the rest of the term.

This is the idea behind *pretesting*, the latest permutation of the testing effect. In a series of experiments, psychologists like Roediger, Karpicke, the Bjorks, and Kornell have found that, in some circumstances, unsuccessful retrieval attempts—i.e., wrong answers—aren't merely random failures. Rather, the attempts themselves alter how we think about, and store, the information contained in the questions. On some kinds of tests, particularly multiple-choice, we learn from answering incorrectly—especially when given the correct answer soon afterward.

That is, *guessing wrongly* increases a person's likelihood of nailing that question, or a related one, on a later test.

That's a sketchy-sounding proposition on its face, it's true. Bombing tests on stuff you don't know sounds more like a recipe for discouragement and failure than an effective learning strategy. The best way to appreciate this is to try it yourself. That means taking another test. It'll be a short one, on something you don't know well—in my case, let's make it the capital cities of African nations. Choose any twelve and have a friend make up a simple multiple-choice quiz, with five possible answers for each nation. Give yourself ten seconds on each question; after each one, have your friend tell you the correct answer.

Ready? Put the smartphone down, close the computer, and give it a shot. Here are a few samples:

BOTSWANA:

- Gaborone
- Dar es Salaam
- Hargeisa

- Oran
- Zaria

(Friend: “Gaborone”)

GHANA:

- Huambo
- Benin
- Accra
- Maputo
- Kumasi

(Friend: “Accra”)

LESOTHO:

- Lusaka
- Juba
- Maseru
- Cotonou
- N’Djamena

(Friend: “Maseru”)

And so on. You’ve just taken a test on which you’ve guessed, if you’re anything like me, mostly wrong. Has taking that test improved your knowledge of those twelve capitals? Of course it has. Your friend *gave you the answers* after each question. Nothing surprising there.

We’re not quite done, though. That was Phase 1 of our experiment, pretesting. Phase 2 will be what we think of as traditional studying. For that, you will need to choose another twelve unfamiliar nations, with the correct answer listed alongside, and then sit down and try to memorize them. Nigeria—Abuja. Eritrea—Asmara. Gambia—Banjul. Take the same amount of time—two minutes—as you took on the multiple-choice test. That’s it. You’re done for the day.

You have now effectively studied the capital cities of twenty-four African nations. You studied the first half by taking a multiple-choice pretest. You studied the other half the old-fashioned way, by straight memorization. We’re going to compare your knowledge of the first twelve to your knowledge of the second twelve.

Tomorrow, take a multiple-choice test on all twenty-four of those nations, also with five possible choices under each nation. When you're done, compare the results. If you're like most people, you scored 10 to 20 percent higher on the countries in that first group, the ones where you guessed before hearing the correct answer. In the jargon of the field, your "unsuccessful retrieval attempts potentiated learning, increasing successful retrieval attempts on subsequent tests."

In plain English: The act of guessing engaged your mind in a different and more demanding way than straight memorization did, deepening the imprint of the correct answers. In even plainer English, the pretest drove home the information in a way that studying-as-usual did not.

Why? No one knows for sure. One possible explanation is that pretesting is another manifestation of desirable difficulty. You work a little harder by guessing first than by studying directly. A second possibility is that the wrong guesses eliminate the fluency illusion, the false impression that you knew the capital of Eritrea because you just saw or studied it. A third is that, in simply memorizing, you saw only the correct answer and weren't thrown off by the other four alternatives—the way you would be on a test. "Let's say you're studying capitals and you see that Australia's is Canberra," Robert Bjork told me. "Okay, 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."

Taking a practice test provides us something else as well—a glimpse of the teacher's hand. "Even when you get wrong answers, it seems to improve subsequent study," Robert Bjork added, "because the test adjusts our thinking in some way to the kind of material we need to know."

That's a good thing, and not just for us. It's in the teacher's interest, 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 organize it, mentally, and use it to make judgments about what's important and what's less so. To Elizabeth Bjork, that seemed the best explanation for why a pretest would promote more effective subsequent studying—it primes students to notice important concepts later on. To find out, she decided to run a pretesting trial in one of her own classes.

Bjork decided to start small, in her Psychology 100B class at UCLA, on research methods. She wouldn't give a comprehensive *prefinal* on the first day of class. "It was a pilot study, really, and I decided to give the pretests for three individual lectures," she said. "The students would take each pretest a day or two before each of those lectures; we wanted to see whether they remembered the material better later."

She and Nicholas Soderstrom, a postdoctoral fellow, designed the three short pretests to have forty questions each, all multiple-choice. They also put together a cumulative exam to be given *after* the three lectures. The crucial question they wanted to answer was: Do students comprehend and retain pretested material better and longer than they do material that's not on a pretest but *is* in the lectures? To answer that, Bjork and Soderstrom did something clever on the final exam. They filled it with two kinds of

questions: those that were related to the pretest questions and those that were not. “If pretesting helps, then students should do better on related questions during a later exam than on material we covered in the lectures but was not pretested,” Bjork said. This is analogous to the African nation test we devised above. The first twelve capitals were “pretested”; the second twelve were not—they were studied in the usual way. By comparing our scores on the first twelve to the second twelve, on a comprehensive test of all twenty-four, we could judge whether pretesting made any difference.

Bjork and Soderstrom would compare students’ scores on pretest-related questions to their scores on non-pretested ones on the cumulative final. The related questions were phrased differently but often had some of the same possible answers. For example, here’s a pair of related questions, one from the pretest and the next from the cumulative exam:

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.

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 each pretest. Then they attended the relevant lecture a day or two later—in effect, getting the correct answers to the questions they’d just tried to answer. Pretesting is most helpful when people get prompt feedback (just as we did on our African capitals test).

Did those bombed tests make any difference in what the students remembered later? The cumulative exam, covering all three pretested lectures, would tell. Bjork and Soderstrom gave that exam two weeks after the last of the three lectures was presented,

and it used the same format as the others: forty multiple-choice questions, each with five possible answers. Again, some of those exam questions were related to pretest ones and others were not. The result? Success. Bjork's Psych 100B class scored about 10 percent higher on the *related* questions than on the unrelated ones. Not a slam dunk, 10 percent—but not bad for a first attempt. “The best way you could say it for now,” she told me, “is that on the basis of preliminary data, 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 a test, she said, they get an opportunity to see the vocabulary used in the coming lectures and get a sense of what kinds of questions and distinctions between concepts are important.

Pretesting is not an entirely new concept. We have all taken practice tests at one time or another as a way of building familiarity—and to questionable effect. Kids have been taking practice SATs for years, just as adults have taken practices MCATs and GMATs and LSATs. Yet the SAT and tests like it are general-knowledge exams, and the practice runs are primarily about reducing anxiety and giving us a feel for format and timing. The research that the Bjorks, Roediger, Kornell, Karpicke and others have done is different. Their testing effect—pre- or post-study—applies to learning the kind of concepts, terms, and vocabulary that form a *specialized* knowledge base, say of introductory chemistry, biblical analysis, or music theory.

In school, testing is still testing. That's not going to change, not fundamentally. What is changing is our appreciation of what a test is. First, thanks to Gates, the Columbia researcher who studied recitation, it appeared to be at least equivalent to additional study: Answering does not only measure what you remember, it increases overall retention. Then, testing proved itself to be *superior* to additional study, in a broad variety of academic topics, and the same is likely true of things like music and dance, practicing from memory. Now we're beginning to understand that some kinds of tests improve later learning—even if we do poorly on them.

Is it possible that one day teachers and professors will give “prefinals” on the first day of class? Hard to say. A prefinal for an intro class in Arabic or Chinese might be a wash, just because the notations and symbols and alphabet are entirely alien. My guess is that prefinals are likely to be much more useful in humanities courses and the social sciences, because in those courses our minds have some scaffolding of language to work with, before making a guess. “At this point, we don't know what the ideal applications of pretesting are,” Robert Bjork told me. “It's still a very new area.”

Besides, in this book we're in the business of discovering what we can do for ourselves, in our own time. Here's what I would say, based on my conversations with the Bjorks, Roediger, and others pushing the limits of retrieval practice: Testing—recitation, self-examination, pretesting, call it what you like—is an enormously powerful technique capable of much more than simply measuring knowledge. It vanquishes the fluency trap that causes so many of us to think that we're poor test takers. It amplifies the value of our study time. And it gives us—in the case of pretesting—a detailed, specific preview of how we should begin to think about approaching a topic.

Testing has brought fear and self-loathing into so many hearts that changing its definition doesn't come easily. There's too much bad blood. Yet one way to do so is to think of the examination as merely one application of testing—one of many. Those applications remind me of what the great Argentine writer Jorge Luis Borges once said about his craft: "Writing long books is a laborious and impoverishing act of foolishness: expanding in five hundred pages an idea that could be perfectly explained in a few minutes. A better procedure is to pretend that those books already exist and to offer a summary, a commentary."

Pretend that the book already exists. Pretend you already know. Pretend you already can play something by Sabicas, that you already inhaled the St. Crispin's Day speech, that you have philosophy logic nailed to the door. Pretend you already are an expert and give a summary, a commentary—pretend and *perform*. That is the soul of self-examination: pretending you're an expert, just to see what you've got. This goes well beyond taking a quick peek at the "summary questions" at the end of the history chapter before reading, though that's a step in the right direction. Self-examination can be done at home. When working on guitar, I learn a few bars of a piece, slowly, painstakingly—then try to play it from memory several times in a row. When reading through a difficult scientific paper, I put it down after a couple times through and try to explain to someone what it says. If there's no one there to listen (or pretend to listen), I say it out loud to myself, trying as hard as I can to quote from the paper its main points. Many teachers have said that you don't really know a topic until you have to *teach* it, until you have to make it clear to someone else. Exactly right. One very effective way to think of self-examination is to say, "Okay, I've studied this stuff; now it's time to tell my brother, or spouse, or teenage daughter what it all means." If necessary, I write it down from memory. As coherently, succinctly, and clearly as I can.

Remember: These apparently simple attempts to communicate what you've learned, to yourself or others, are not merely a form of self-testing, in the conventional sense, but *studying*—the high-octane kind, 20 to 30 percent more powerful than if you continued sitting on your butt, staring at that outline. Better yet, those exercises will dispel the fluency illusion. They'll expose what you don't know, where you're confused, what you've forgotten—and fast.

That's ignorance of the best kind.