2

To Learn, Retrieve

MIKE EBERSOLD GOT CALLED into a hospital emergency room one afternoon late in 2011 to examine a Wisconsin deer hunter who'd been found lying unconscious in a cornfield. The man had blood at the back of his head, and the men who'd found and brought him in supposed he'd maybe stumbled and cracked his skull on something.

Ebersold is a neurosurgeon. The injury had brain protruding, and he recognized it as a gunshot wound. The hunter regained consciousness in the ER, but when asked how he'd hurt himself, he had no idea.

Recounting the incident later, Ebersold said, "Somebody from some distance away must have fired what appeared to be a 12-gauge shotgun, which arced over God only knows what distance, hit this guy in the back of his head, fractured his skull, and lodged into the brain about an inch. It must have been pretty much spent, or it would have gone deeper."[1]

Ebersold is tall, slender, and counts among his forebears the Dakota chiefs named Wapasha and the French fur traders named Rocque who populated this part of the Mississippi River Valley where the Mayo brothers would later found their famous clinic. Ebersold's formal training included four years of college, four years of medical school, and seven years of neurosurgery training—building a foundation of knowledge and skills that has been broadened and deepened through continuing medical education classes, consultations with his colleagues, and his practice at the Mayo Clinic and elsewhere. He carries himself with a midwestern modesty that belies a career that counts a long list of high-profile patients who have sought out his services. When President Ronald Reagan needed treatment for injuries after a fall from his horse, Ebersold participated in the surgery and postsurgical care. When Sheikh Zayed bin Sultan Al Nahyan, president of the United Arab Emirates, needed delicate spinal repair, he and what seemed like half the nation's ministry and security forces settled in Rochester while Mike Ebersold made the repair and oversaw Zayed's recovery. Following a long career at Mayo, Mike had returned to help out at the clinic in Wisconsin, feeling indebted to it for his early medical training. The hunter whose bad luck put him in the way of an errant 12-gauge slug was luckier than he likely knows that Mike was on the job that day.

The bullet had entered an area of the skull beneath which there is a large venous sinus, a soft-tissue channel that drains the brain cavity. As he examined the hunter, Ebersold knew from experience that when he opened up the wound, there was a high probability he would find this vein was torn. As he described it,

> You say to yourself, "This patient is going to need surgery. There's brain coming out of the wound. We have to clean this

up and repair this as best we can, but in so doing we may get into this big vein and that could be very, very serious." So you go through the checklist. You say, "I might need a blood transfusion for this patient," so you set up some blood. You review the steps, A, B, C, and D. You set up the operating room, telling them ahead of time what you might be encountering. All of this is sort of protocol, pretty much like a cop getting ready to pull over a car, you know what the book says, you've gone through all these steps.

Then you get to the operating room, and now you're still in this mode where you have time to think through it. You say, "Gee, I don't want to just go and pull that bullet out if there might be major bleeding. What I'll try to do is I'll work around the edges and get things freed up so I'm ready for what could go wrong, and then I'll pull it out."

It turned out that the bullet and bone were lodged in the vein, serving as plugs, another lucky turn for the hunter. If the wound hadn't corked itself in the field, he would not have lived for more than two or three minutes. When Ebersold removed the bullet, the fractured bone chips fell away, and the vein let loose in a torrent. "Within five minutes, you've lost two or so units of blood and now you sort of transfer out of the mode where you're thinking through this, going through the options. Now it becomes reflex, mechanical. You know it's going to bleed very, very much, so you have a very short time. You're just thinking, 'I have to get a suture around this structure, and I know from previous experience I have to do it in this particular way.' "

The vein in question, which is about the size of an adult's small finger, was torn in several places over a distance of about an inch and a half. It needed to be tied off above and below the rupture, but it's a flat structure that he knows well: you

can't just put a stitch around it, because when you tighten it, the tissue tears, and the ligature leaks. Working urgently and mechanically, he fell back on a technique he'd developed out of necessity in past surgeries involving this vein. He cut two little pieces of muscle, from where the patient's skin had been opened up in surgery, and imported them to the site and stitched the ends of the torn vein to them. These plugs of muscle served to close the vein without deflecting its natural shape or tearing its tissue. It's a solution Mike has taught himself—one he says you won't find written anywhere, but handy in the moment, to say the least. In the sixty or so seconds it took to do, the patient lost another two hundred cubic centimeters of blood, but once the plugs were in place, the bleeding stopped. "Some people can't tolerate this sinus vein being closed off. They get increased brain pressure because the blood doesn't drain properly. But this patient was one of the fortunate who can." The hunter left the hospital a week later. He was minus some peripheral vision but otherwise remarkably unscathed from a very close brush with mortality.

Reflection Is a Form of Practice

What inferences can we draw from this story about how we learn and remember? In neurosurgery (and, arguably, in all aspects of life from the moment you leave the womb), there's an essential kind of learning that comes from reflection on personal experience. Ebersold described it this way:

> A lot of times something would come up in surgery that I had difficulty with, and then I'd go home that night thinking about what happened and what could I do, for example, to improve the way a suturing went. How can I take a bigger bite with my needle, or a smaller bite, or should the stitches be closer together? What if I modified it this way or that way? Then the

> next day back, I'd try that and see if it worked better. Or even if it wasn't the next day, at least I've thought through this, and in so doing I've not only revisited things that I learned from lectures or from watching others performing surgery but also I've complemented that by adding something of my own to it that I missed during the teaching process.

Reflection can involve several cognitive activities that lead to stronger learning: retrieving knowledge and earlier training from memory, connecting these to new experiences, and visualizing and mentally rehearsing what you might do differently next time.

It was this kind of reflection that originally had led Ebersold to try a new technique for repairing the sinus vein at the back of the head, a technique he practiced in his mind and in the operating room until it became the kind of reflexive maneuver you can depend on when your patient is spouting blood at two hundred cubic centimeters a minute.

To make sure the new learning is available when it's needed, Ebersold points out, "you memorize the list of things that you need to worry about in a given situation: steps A, B, C, and D," and you drill on them. Then there comes a time when you get into a tight situation and it's no longer a matter of thinking through the steps, it's a matter of reflexively taking the correct action. "Unless you keep recalling this maneuver, it will not become a reflex. Like a race car driver in a tight situation or a quarterback dodging a tackle, you've got to act out of reflex before you've even had time to think. Recalling it over and over, practicing it over and over. That's just so important."

The Testing Effect

A child stringing cranberries on a thread goes to hang them on the tree, only to find they've slipped off the other end. Without the knot, there's no making a string. Without the knot there's no necklace, there's no beaded purse, no magnificent tapestry. Retrieval ties the knot for memory. Repeated retrieval snugs it up and adds a loop to make it fast.

Since as far back as 1885, psychologists have been plotting "forgetting curves" that illustrate just how fast our cranberries slip off the string. In very short order we lose something like 70 percent of what we've just heard or read. After that, forgetting begins to slow, and the last 30 percent or so falls away more slowly, but the lesson is clear: a central challenge to improving the way we learn is finding a way to interrupt the process of forgetting.[2]

The power of retrieval as a learning tool is known among psychologists as the testing effect. In its most common form, testing is used to measure learning and assign grades in school, but we've long known that the act of retrieving knowledge from memory has the effect of making that knowledge easier to call up again in the future. In his essay on memory, Aristotle wrote: "exercise in repeatedly recalling a thing strengthens the memory." Francis Bacon wrote about this phenomenon, as did the psychologist William James. Today, we know from empirical research that practicing retrieval makes learning stick far better than reexposure to the original material does. This is the testing effect, also known as the retrieval-practice effect.[3]

To be most effective, retrieval must be repeated again and again, in spaced out sessions so that the recall, rather than becoming a mindless recitation, requires some cognitive effort. Repeated recall appears to help memory consolidate into a cohesive representation in the brain and to strengthen and

multiply the neural routes by which the knowledge can later be retrieved. In recent decades, studies have confirmed what Mike Ebersold and every seasoned quarterback, jet pilot, and teenaged texter knows from experience—that repeated retrieval can so embed knowledge and skills that they become reflexive: the brain acts before the mind has time to think.

Yet despite what research and personal experience tell us about the power of testing as a learning tool, teachers and students in traditional educational settings rarely use it as such, and the technique remains little understood or utilized by teachers or students as a learning tool in traditional educational settings. Far from it.

In 2010 the *New York Times* reported on a scientific study that showed that students who read a passage of text and then took a test asking them to recall what they had read retained an astonishing 50 percent more of the information a week later than students who had not been tested. This would seem like good news, but here's how it was greeted in many online comments:

> "Once again, another author confuses learning with recalling information."
>
> "I personally would like to avoid as many tests as possible, especially with my grade on the line. Trying to learn in a stressful environment is no way to help retain information."
>
> "Nobody should care whether memorization is enhanced by practice testing or not. Our children cannot *do* much of anything anymore."[4]

Forget memorization, many commenters argued; education should be about high-order skills. Hmmm. If memorization is irrelevant to complex problem solving, don't tell your

neurosurgeon. The frustration many people feel toward standardized, "dipstick" tests given for the sole purpose of measuring learning is understandable, but it steers us away from appreciating one of the most potent learning tools available to us. Pitting the learning of basic knowledge against the development of creative thinking is a false choice. Both need to be cultivated. The stronger one's knowledge about the subject at hand, the more nuanced one's creativity can be in addressing a new problem. Just as knowledge amounts to little without the exercise of ingenuity and imagination, creativity absent a sturdy foundation of knowledge builds a shaky house.

Studying the Testing Effect in the Lab

The testing effect has a solid pedigree in empirical research. The first large-scale investigation was published in 1917. Children in grades 3, 5, 6, and 8 studied brief biographies from *Who's Who in America*. Some of them were directed to spend varying lengths of the study time looking up from the material and silently reciting to themselves what it contained. Those who did not do so simply continued to reread the material. At the end of the period, all the children were asked to write down what they could remember. The recall test was repeated three to four hours later. All the groups who had engaged in the recitation showed better retention than those who had not done so but had merely continued to review the material. The best results were from those spending about 60 percent of the study time in recitation.

A second landmark study, published in 1939, tested over three thousand sixth graders across Iowa. The kids studied six-hundred-word articles and then took tests at various times before a final test two months later. The experiment showed a couple of interesting results: the longer the first test was delayed, the greater the forgetting, and second, once a student

had taken a test, the forgetting nearly stopped, and the student's score on subsequent tests dropped very little.[5]

Around 1940, interest turned to the study of forgetting, and investigating the potential of testing as a form of retrieval practice and as a learning tool fell out of favor. So did the use of testing as a research tool: since testing interrupts forgetting, you can't use it to measure forgetting because that "contaminates" the subject.

Interest in the testing effect resurfaced in 1967 with the publication of a study showing that research subjects who were presented with lists of thirty-six words learned as much from repeated testing after initial exposure to the words as they did from repeated studying. These results—that testing led to as much learning as studying did—challenged the received wisdom, turned researchers' attention back to the potential of testing as a learning tool, and stimulated a boomlet in testing research.

In 1978, researchers found that massed studying (cramming) leads to higher scores on an immediate test but results in faster forgetting compared to practicing retrieval. In a second test two days after an initial test, the crammers had forgotten 50 percent of what they had been able to recall on the initial test, while those who had spent the same period practicing retrieval instead of studying had forgotten only 13 percent of the information recalled initially.

A subsequent study was aimed at understanding what effect taking multiple tests would have on subjects' long-term retention. Students heard a story that named sixty concrete objects. Those students who were tested immediately after exposure recalled 53 percent of the objects on this initial test but only 39 percent a week later. On the other hand, a group of students who learned the same material but were not tested at all until a week later recalled 28 percent. Thus, taking a single test boosted performance by 11 percent after a week.

But what effect would three immediate tests have relative to one? Another group of students were tested three times after initial exposure and a week later they were able to recall 53 percent of the objects—the same as on the initial test for the group receiving one test. In effect, the group that received three tests had been "immunized" against forgetting, compared to the one-test group, and the one-test group remembered more than those who had received no test immediately following exposure. Thus, and in agreement with later research, multiple sessions of retrieval practice are generally better than one, especially if the test sessions are spaced out.[6]

In another study, researchers showed that simply asking a subject to fill in a word's missing letters resulted in better memory of the word. Consider a list of word pairs. For a pair like *foot-shoe*, those who studied the pair intact had lower subsequent recall than those who studied the pair from a clue as obvious as *foot-s_ _e*. This experiment was a demonstration of what researchers call the "generation effect." The modest effort required to generate the cued answer while studying the pairs strengthened memory of the target word tested later (shoe). Interestingly, this study found that the ability to recall the word pair on later tests was greater if the practice retrieval was delayed by twenty intervening word pairs than when it came immediately after first studying the pair.[7] Why would that be? One argument suggested that the greater effort required by the delayed recall solidified the memory better. Researchers began to ask whether the schedule of testing mattered.

The answer is yes. When retrieval practice is spaced, allowing some forgetting to occur between tests, it leads to stronger long-term retention than when it is massed.

Researchers began looking for opportunities to take their inquiries out of the lab and into the classroom, using the kinds of materials students are required to learn in school.

Studying the Testing Effect "In the Wild"

In 2005, we and our colleagues approached Roger Chamberlain, the principal of a middle school in nearby Columbia, Illinois, with a proposition. The positive effects of retrieval practice had been demonstrated many times in controlled laboratory settings but rarely in a regular classroom setting. Would the principal, teachers, kids, and parents of Columbia Middle School be willing subjects in a study to see how the testing effect would work "in the wild"?

Chamberlain had concerns. If this was just about memorization, he wasn't especially interested. His aim is to raise the school's students to higher forms of learning—analysis, synthesis, and application, as he put it. And he was concerned about his teachers, an energetic faculty with curricula and varied instructional methods he was loath to disrupt. On the other hand, the study's results could be instructive, and participation would bring enticements in the form of smart boards and "clickers"—automated response systems—for the classrooms of participating teachers. Money for new technology is famously tight.

A sixth grade social studies teacher, Patrice Bain, was eager to give it a try. For the researchers, a chance to work in the classroom was compelling, and the school's terms were accepted: the study would be minimally intrusive by fitting within existing curricula, lesson plans, test formats, and teaching methods. The same textbooks would be used. The only difference in the class would be the introduction of occasional short quizzes. The study would run for three semesters (a year and a half), through several chapters of the social studies textbook, covering topics such as ancient Egypt, Mesopotamia, India, and China. The project was launched in 2006. It would prove to be a good decision.

For the six social studies classes a research assistant, Pooja Agarwal, designed a series of quizzes that would test students on roughly one-third of the material covered by the teacher. These quizzes were for "no stakes," meaning that scores were not counted toward a grade. The teacher excused herself from the classroom for each quiz so as to remain unaware of which material was being tested. One quiz was given at the start of class, on material from assigned reading that hadn't yet been discussed. A second was given at the end of class after the teacher had covered the material for the day's lesson. And a review quiz was given twenty-four hours before each unit exam.

There was concern that if students tested better in the final exam on material that had been quizzed than on material not quizzed, it could be argued that the simple act of reexposing them to the material in the quizzes was responsible for the superior learning, not the retrieval practice. To counter this possibility, some of the nonquizzed material was interspersed with the quiz material, provided as simple review statements, like "The Nile River has two major tributaries: the White Nile and the Blue Nile," with no retrieval required. The facts were quizzed for some classes but just restudied for others.

The quizzes took only a few minutes of classroom time. After the teacher stepped out of the room, Agarwal projected a series of slides onto the board at the front of the room and read them to the students. Each slide presented either a multiple choice question or a statement of fact. When the slide contained a question, students used clickers (handheld, cell-phone-like remotes) to indicate their answer choice: A, B, C, or D. When all had responded, the correct answer was revealed, so as to provide feedback and correct errors. (Although teachers were not present for these quizzes, under normal circumstances,

with teachers administering quizzes, they would see immediately how well students are tracking the study material and use the results to guide further discussion or study.)

Unit exams were the normal pencil-and-paper tests given by the teacher. Exams were also given at the end of the semester and at the end of the year. Students had been exposed to all of the material tested in these exams through the teacher's normal classroom lessons, homework, worksheets, and so on, but they had also been quizzed three times on one-third of the material, and they had seen another third presented for additional study three times. The balance of the material was neither quizzed nor additionally reviewed in class beyond the initial lesson and whatever reading a student may have done.

The results were compelling: The kids scored a full grade level higher on the material that had been quizzed than on the material that had not been quizzed. Moreover, test results for the material that had been reviewed as statements of fact but not quizzed were no better than those for the nonreviewed material. Again, mere rereading does not much help.

In 2007, the research was extended to eighth grade science classes, covering genetics, evolution, and anatomy. The regimen was the same, and the results equally impressive. At the end of three semesters, the eighth graders averaged 79 percent (C+) on the science material that had not been quizzed, compared to 92 percent (A–) on the material that had been quizzed.

The testing effect persisted eight months later at the end-of-year exams, confirming what many laboratory studies have shown about the long-term benefits of retrieval practice. The effect doubtless would have been greater if the retrieval practice had continued and occurred once a month, say, in the intervening months.[8]

The lesson from these studies has been taken to heart by many of the teachers at Columbia Middle School. Long after concluding their participation in the research studies, Patrice Bain's sixth grade social studies classes continue today to follow a schedule of quizzes before lessons, quizzes after lessons, and then a review quiz prior to the chapter test. Jon Wehrenberg, an eighth grade history teacher who was not part of the research, has knitted retrieval practice into his classroom in many different forms, including quizzing, and he provides additional online tools at his website, like flashcards and games. After reading passages on the history of slavery, for example, his students are asked to write down ten facts about slavery they hadn't known before reading the passages. You don't need electronic gadgetry to practice retrieval.

Seven sixth and seventh graders needing to improve their reading and comprehension skills sat in Michelle Spivey's English classroom one period recently with their reading books open to an amusing story. Each student was invited to read a paragraph aloud. Where a student stumbled, Miss Spivey had him try again. When he'd gotten it right, she probed the class to explain the meaning of the passage and what might have been going on in the characters' minds. Retrieval and elaboration; again, no technology required.

Quizzes at Columbia Middle School are not onerous events. Following completion of the research studies, students' views were surveyed on this question. Sixty-four percent said the quizzing reduced their anxiety over unit exams, and 89 percent felt it increased learning. The kids expressed disappointment on days when clickers were not used, because the activity broke up the teacher's lecture and proved enjoyable.

Principal Chamberlain, when asked what he thought the study results indicated, replied simply: "Retrieval practice has a significant impact on kids' learning. This is telling us that

it's valuable, and that teachers are well advised to incorporate it into their instructional technique."[9]

Are similar effects found at a later age?

Andrew Sobel teaches a class in international political economics at Washington University in St. Louis, a lecture course populated by 160–170 students, mostly freshmen and sophomores. Over a period of several years he noticed a growing problem with attendance. On any given day by midsemester, 25–35 percent of the class would be absent, compared to earlier in the semester when maybe 10 percent would be absent. The problem wasn't unique to his class, he says. A lot of professors give students their PowerPoint slides, so the students just stop coming to class. Sobel fought back by withholding his slides, but by the end of the semester, many students stopped showing up anyway. The class syllabus included two big tests, a midterm and a final. Looking for some way to leverage attendance, Sobel replaced the big tests with nine pop quizzes. Because the quizzes would determine the course grade and would be unannounced, students would be well advised to show up for class.

The results were distressing. Over the semester, a third or more of the students bailed out. "I really got hammered in the teaching reviews," Sobel told us. "The kids hated it. If they didn't do well on a quiz they dropped the course rather than get a bad grade in it. Of those who stayed, I got this bifurcation between those who actually showed up and did the work, and those who didn't. I found myself handing out A-plusses, which I'd never given before, and more Cs than I'd ever given."[10]

With so much pushback, he had little choice but to drop the experiment and reinstate the old format, lectures with a midterm and final. A couple of years later, however, after hearing a

presentation about the learning benefits of testing, he added a third major test during the semester to see what effect it might have on his students' learning. They did better, but not by as much as he'd hoped, and the attendance problems persisted.

He scratched his head and changed the syllabus once again. This time he announced that there would be nine quizzes during the semester, and he was explicit about when they would be. No surprises, and no midterm or final exams, because he didn't want to give up that much of his lecture time.

Despite fears that enrollments would plummet again, they actually increased by a handful. "Unlike the pop quizzes, which kids hate, these were all on the syllabus. If they missed one it was their own fault. It wasn't because I surprised them or was being pernicious. They were comfortable with that." Sobel took satisfaction in seeing attendance improve as well. "They would skip some classes on the days they didn't have a quiz, particularly the spring semester, but they showed up for the quizzes."

Like the course, the quizzes were cumulative, and the questions were similar to those on the exams he used to give, but the quality of the answers he was getting by midsemester was much better than he was accustomed to seeing on the midterms. Five years into this new format, he's sold on it. "The quality of discussions in class has gone way up. I see that big a difference in their written work, just by going from three exams to nine quizzes." By the end of the semester he has them writing paragraphs on the concepts covered in class, sometimes a full-page essay, and the quality is comparable to what he's seeing in his upper division classes.

"Anybody can design this structure. But I also realize that, Oh, god, if I'd done this years ago I would have taught them that much more stuff. The interesting thing about adopting this strategy is I now recognize that as good a teacher as I

might think I am, my teaching is only a component of their learning, and how I structure it has a lot to do with it, maybe even more." Meanwhile, the course enrollment has grown to 185 and counting.

Exploring Nuances

Andy Sobel's example is anecdotal and likely reflects a variety of beneficial influences, not least being the cumulative learning effects that accrue like compounded interest when course material is carried forward in a regime of quizzes across an entire semester. Nonetheless, his experience squares with empirical research designed to tease apart the effects and nuances of testing.

For example, in one experiment college students studied prose passages on various scientific topics like those taught in college and then either took an immediate recall test after the initial exposure or restudied the material. After a delay of two days, the students who took the initial test recalled more of the material than those who simply restudied it (68 v. 54 percent), and this advantage was sustained a week later (56 v. 42 percent). Another experiment found that after one week a study-only group showed the most forgetting of what they initially had been able to recall, forgetting 52 percent, compared to a repeated-testing group, who forgot only 10 percent.[11]

How does giving feedback on wrong answers to test questions affect learning? Studies show that giving feedback strengthens retention more than testing alone does, and, interestingly, some evidence shows that delaying the feedback briefly produces better long-term learning than immediate feedback. This finding is counterintuitive but is consistent with researchers'

discoveries about how we learn motor tasks, like making layups or driving a golf ball toward a distant green. In motor learning, trial and error with delayed feedback is a more awkward but effective way of acquiring a skill than trial and correction through immediate feedback; immediate feedback is like the training wheels on a bicycle: the learner quickly comes to depend on the continued presence of the correction.

In the case of learning motor skills, one theory holds that when there's immediate feedback it comes to be part of the task, so that later, in a real-world setting, its absence becomes a gap in the established pattern that disrupts performance. Another idea holds that frequent interruptions for feedback make the learning sessions too variable, preventing establishment of a stabilized pattern of performance.[12]

In the classroom, delayed feedback also yields better long-term learning than immediate feedback does. In the case of the students studying prose passages on science topics, some were shown the passage again even while they were asked to answer questions about it, in effect providing them with continuous feedback during the test, analogous to an open-book exam. The other group took the test without the study material at hand and only afterward were given the passage and instructed to look over their responses. Of course, the open-book group performed best on the immediate test, but those who got corrective feedback after completing the test retained the learning better on a later test. Delayed feedback on written tests may help because it gives the student practice that's spaced out in time; as discussed in the next chapter, spacing practice improves retention.[13]

Are some kinds of retrieval practice more effective for long-term learning than others? Tests that require the learner to

supply the answer, like an essay or short-answer test, or simply practice with flashcards, appear to be more effective than simple recognition tests like multiple choice or true/false tests. However, even multiple choice tests like those used at Columbia Middle School can yield strong benefits. While any kind of retrieval practice generally benefits learning, the implication seems to be that where more cognitive effort is required for retrieval, greater retention results. Retrieval practice has been studied extensively in recent years, and an analysis of these studies shows that even a single test in a class can produce a large improvement in final exam scores, and gains in learning continue to increase as the number of tests increases.[14]

Whichever theories science eventually tells us are correct about *how* repeated retrieval strengthens memory, empirical research shows us that the testing effect is real—that the act of retrieving a memory changes the memory, making it easier to retrieve again later.

How widely is retrieval practice used as a study technique? In one survey, college students were largely unaware of its effectiveness. In another survey, only 11 percent of college students said they use this study strategy. Even when they did report testing themselves, they mostly said they did it to discover what they didn't know, so they could study that material more. That's a perfectly valid use of testing, but few students realize that retrieval itself creates greater retention.[15]

Is repeated testing simply a way to expedite rote learning? In fact, research indicates that testing, compared to rereading, can facilitate better transfer of knowledge to new contexts and problems, and that it improves one's ability to retain and

retrieve material that is related but not tested. Further research is needed on this point, but it seems that retrieval practice can make information more accessible when it is needed in various contexts.

Do students resist testing as a tool for learning? Students do generally dislike the idea of tests, and it's not hard to see why, in particular in the case of high-stakes tests like midterms and finals, where the score comes with significant consequences. Yet in all studies of testing that reported students' attitudes, the students who were tested frequently rated their classes more favorably at the end of the semester than those tested less frequently. Those who were frequently tested reached the end of the semester on top of the material and did not need to cram for exams.

How does taking a test affect subsequent studying? After a test, students spend more time restudying the material they missed, and they learn more from it than do their peers who restudy the material without having been tested. Students whose study strategies emphasize rereading but not self-testing show overconfidence in their mastery. Students who have been quizzed have a double advantage over those who have not: a more accurate sense of what they know and don't know, and the strengthening of learning that accrues from retrieval practice.[16]

Are there any further, indirect benefits of regular, low-stakes classroom testing? Besides strengthening learning and retention, a regime of this kind of testing improves student attendance. It increases studying before class (because students

know they'll be quizzed), increases attentiveness during class if students are tested at the end of class, and enables students to better calibrate what they know and where they need to bone up. It's an antidote to mistaking fluency with the text, resulting from repeated readings, for mastery of the subject. Frequent low-stakes testing helps dial down test anxiety among students by diversifying the consequences over a much larger sample: no single test is a make-or-break event. And this kind of testing enables instructors to identify gaps in students' understanding and adapt their instruction to fill them. These benefits of low-stakes testing accrue whether instruction is delivered online or in the classroom.[17]

The Takeaway

Practice at retrieving new knowledge or skill from memory is a potent tool for learning and durable retention. This is true for anything the brain is asked to remember and call up again in the future—facts, complex concepts, problem-solving techniques, motor skills.

Effortful retrieval makes for stronger learning and retention. We're easily seduced into believing that learning is better when it's easier, but the research shows the opposite: when the mind has to work, learning sticks better. The greater the effort to retrieve learning, provided that you succeed, the more that learning is strengthened by retrieval. After an initial test, *delaying subsequent retrieval practice* is more potent for reinforcing retention than immediate practice, because delayed retrieval requires more effort.

Repeated retrieval not only makes memories more durable but produces knowledge that can be retrieved more readily, in more varied settings, and applied to a wider variety of problems.

While cramming can produce better scores on an immediate exam, the advantage quickly fades because there is much greater forgetting after rereading than after retrieval practice. The benefits of retrieval practice are *long-term*.

Simply including one test (retrieval practice) in a class yields a large improvement in final exam scores, and gains continue to increase as the frequency of classroom testing increases.

Testing doesn't need to be initiated by the instructor. Students can practice retrieval anywhere; no quizzes in the classroom are necessary. Think flashcards—the way second graders learn the multiplication tables can work just as well for learners at any age to quiz themselves on anatomy, mathematics, or law. Self-testing may be unappealing because it takes more effort than rereading, but as noted already, the greater the effort at retrieval, the more will be retained.

Students who take practice tests have *a better grasp of their progress* than those who simply reread the material. Similarly, such testing enables an instructor to *spot gaps and misconceptions* and adapt instruction to correct them.

Giving students *corrective feedback* after tests keeps them from incorrectly retaining material they have misunderstood and produces better learning of the correct answers.

Students in classes that incorporate low-stakes quizzing come to embrace the practice. Students who are tested frequently rate their classes more favorably.

What about Principal Roger Chamberlain's initial concerns about practice quizzing at Columbia Middle School—that it might be nothing more than a glorified path to rote learning?

When we asked him this question after the study was completed, he paused for a moment to gather his thoughts. "What

I've really gained a comfort level with is this: for kids to be able to evaluate, synthesize, and apply a concept in different settings, they're going to be much more efficient at getting there when they have the base of knowledge and the retention, so they're not wasting time trying to go back and figure out what that word might mean or what that concept was about. It allows them to go to a higher level."