

REVIEW

Using Testing as a Learning Tool

Brenda W. Yang, BS,^a Juan Razo, BS,^b Adam M. Persky, PhD^{b,c}

^a Duke University, Psychology and Neuroscience, Center for Cognitive Neuroscience, Durham, North Carolina

^b University of North Carolina at Chapel Hill, Eshelman School of Pharmacy, Chapel Hill, North Carolina

^c Associate Editor, *American Journal of Pharmaceutical Education*, Arlington, Virginia

Submitted August 9, 2018; accepted April 10, 2019; published November 2019.

Objective. The purpose of this review is to discuss some principles from cognitive psychology regarding the benefits of testing and translate those findings into practical applications for instruction and studying.

Findings. Testing or retrieval practice is superior to re-study for promoting long-term retention. The benefits of testing can be seen with open-ended responses (eg, cued or free recall) and multiple choice questions. The use of multiple-choice questions during testing may have an additional benefit as it may stabilize information that is stored in memory but temporarily inaccessible due to disuse (eg, marginal knowledge).

Summary. Testing can have multiple learning benefits. We emphasize that incorporating opportunities for retrieval after teaching is an essential component of lasting learning. In addition, retrieval practice can be incorporated in all aspects of instruction.

Keywords: formative assessment, summative assessment, multiple-choice questions, testing effect, retrieval

INTRODUCTION

What role do assessments play in the classroom? Educators typically consider examinations as summative (evaluating work and assigning grades) or formative (to gain information about student learning and to provide feedback to students). However, tests are powerful tools to promote lasting learning in their own right. The purpose of this review is both practical and theoretical. We provide concrete strategies to help instructors design curricula and give advice to students on how to study. However, no single list of recommendations will be optimal for every classroom. Thus, this review situates suggestions for the classroom within a body of knowledge from cognitive psychology providing empirical justification and theory for how students learn and remember. The idea is that instructors grounded in the reasoning behind recommendations will be better empowered to effectively implement assessments suited to their own situation.

We begin with a summary of the psychology of learning and memory. We then provide an overview of considerations in implementing testing as learning in the

classrooms, such as format and timing, before discussing the potential benefits of multiple-choice testing in particular. We conclude with recommendations for educators designing a course or curriculum, study strategies for students, and an illustrative vignette of some of these recommendations implemented within a health science course.

The Psychology of Learning. What is learning? While students and instructors often have their own answers, the present discussion uses a psychological lens, which partitions the process into steps to allow for better study of each component (Figure 1).

Starting from the left of Figure 1, sensory input in a classroom context might consist of an instructor's lecture or a patient presentation's visual content. When a student attends to these stimuli, a representation of this information enters the student's working memory (the mind's workbench) where information is consciously held and can be manipulated. However, the capacity of working memory is relatively small, and information here fades quickly (eg, attempting complex mental arithmetic is an easy way to experience the limited nature of working memory). Much of a student's daily experience enters his memory only to be lost, such as what was on the radio a month ago, which pair of socks were worn last Saturday, or the content of a class lecture from even a few days ago. While information can be maintained in working memory by simple repetition, such as repeating a phone number to

Corresponding Author: Adam M. Persky, University of North Carolina at Chapel Hill, UNC Eshelman School of Pharmacy, 2312 Kerr Hall, CB#7569, Chapel Hill, NC 27599. Tel: 919-966-9104. Fax: 919-962-0644. Email: apersky@unc.edu.

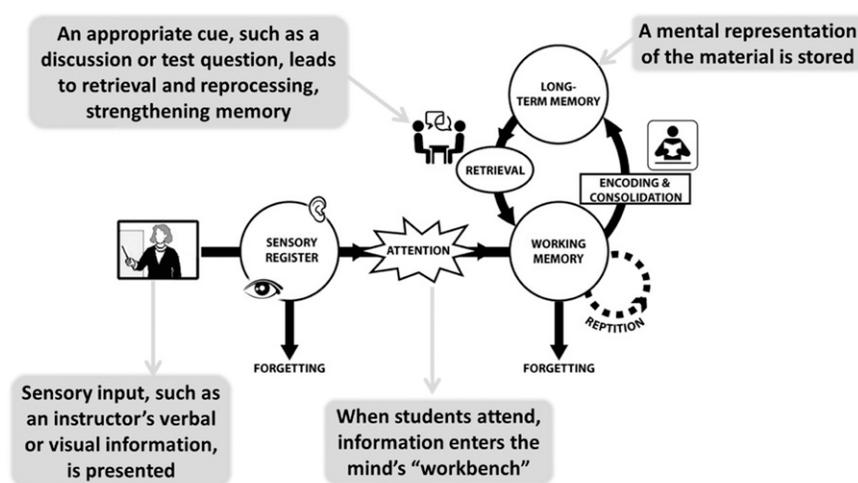


Figure 1. A Review of Memory Formation, Highlighting Major Elements of Modern Models of Memory^{2,66,67}

hold it in mind while looking for a pen and paper, this approach is clearly not ideal for most situations. To be available later, mnemonic information must be *encoded* or stored into “long-term memory,” an umbrella term for anything that can be remembered for longer than a few minutes. In contrast to working memory, the capacity of long-term memory is extraordinarily large and functionally unlimited, holding information as diverse as the capital of France, how to start a car, what a camping trip was like, and in what circumstances to maximize the dose of hydrochlorothiazide or add an ACE-inhibitor to a hypertensive patient’s drug regimen.¹ In healthy brains, information in long-term memory is likely stored indefinitely.^{2,3}

Instructional effort is often lavished on how students initially encode information. However, what occurs after encoding is just as critical, if not more so, for successful learning. Representations of memories in the brain are far from stable; rather, they are susceptible to reorganization in a set of processes called *consolidation*.⁴⁻⁸ Such processes can involve replaying the experience, assigning it meaning, reinterpreting information, or making connections. After encoding and any consolidation processes that may occur, memory traces can be *retrieved* or recalled at some later time. While retrieval can happen involuntarily (ie, popping into a person’s mind), memories are typically remembered from a *cue*, which functions as a hook to extract information. For example, a student might retrieve the response “hypokalemia” from their memory in response to a teacher’s question about the major side effects of hydrochlorothiazide.

If the capacity for remembering is virtually infinite and memories are stored for a lifetime, why do students labor over flashcards to retain information and educators bemoan how little information students retain from

previous courses? In short, why are there so many errors in human memory? One critical reason is that not all available memories are accessible at a given moment in time.^{9,10} The feeling of not being able to access a piece of information is particularly palpable in “tip-of-the-tongue” states, as people tend to have an accurate assessment about whether they know something without being able to retrieve it in that moment.¹¹ How easily something is retrieved depends on a variety of factors, including how many cues are linked to the material and how recently and often the information has been retrieved. In general, while stored memories are long-lasting, retrieval of them is best described as “erratic, highly fallible, and heavily cue-dependent.”¹² Even information that was very well learned and often rehearsed (such as one’s childhood phone number) quickly degrades without regular practice retrieving it. Given this background, we define learning as both acquiring knowledge and skills as well as having them readily available from memory so the individual can make sense of future problems and opportunities.¹³

Testing as Learning. When considering learning as described above, the ways in which learners retrieve information is as critical, if not more so, than how they were initially exposed to it. Testing is an invaluable opportunity for learning, in addition to its more commonly considered roles in evaluating student learning (ie, as summative assessments) and providing feedback to guide future learning (ie, as formative assessments). Thus, we argue that educators should elevate the role of testing in their course curricula, planning testing opportunities with an eye towards the potential of tests to spur learning. While we use the term “testing” in the present manuscript for simplicity and cohesion with existing research, a more accurate description of what we are referring to might be “retrieval practice.” Testing for learning is not necessarily

evaluative, and using alternate descriptions (eg, practice, quiz) might assuage some of the negative affect towards the term “testing” itself. And while the principles of retrieval practice can be powerfully implemented within the classroom, they can also be operationalized as study strategies eg, flash cards or practice examinations.

The benefits of testing are well-established in controlled laboratory studies as well as classrooms, ranging from primary education to professional schools and across a wide range of content areas. This benefit is typically established by comparing a group of people who are exposed to information and then tested on it (study-test) to a group of people who are exposed to the information repeatedly (study-restudy). Variations on this design include expanding the number of times test and study phases happen, eg, study-test-test-test vs study-restudy-restudy-restudy. Note that the study-restudy control is a fairly robust control group, as rereading is a common learning strategy among students and serves as a more active comparison than a study-nothing group.¹⁴ People in study-test groups consistently perform better on subsequent tasks, even without feedback on performance during the test phase (effect size, $d=0.67$).¹⁵ The benefits of testing are even larger for longer retention intervals (<1 day, $d=0.58$; >1 day, $d=0.78$).^{15,16} Examples of studies demonstrating the “testing effect” conducted with professional students in medical careers are provided in Table 1.

What drives the testing effect? One reason is that it fits a general principle suggesting that the learning process should have some level of difficulty in order to be remembered well.¹⁷ Many strategies that initially slow down learning and make it feel more laborious have a beneficial outcome in the long run. Testing is a canonical example of such a technique.¹⁸ While repeated studying improves recall in the short term (eg, five minutes after studying), testing improves recall over longer intervals

(eg, after a week).¹⁹ Other study techniques show a similar pattern, such as spacing studying out over time and interleaving different topics in practice rather than blocking them. Meta-analyses of psychological literature support the idea that it is the relative difficulty of retrieval, compared to re-study, which partly underpins why testing helps with learning, and especially with retention over time.¹⁵

This finding helps to explain why students tend to choose study strategies not conducive to long-term learning: without the insight of empirical studies, students rely on their subjective experience and intuition to choose strategies that improve short-term performance.¹⁸ Indeed, students who only re-study material predict they will recall information better in a week than those who are tested on it, although empirical results demonstrate exactly the reverse.^{19,20} Moreover, rereading textbooks or class notes feels easier than retrieval practice, as the added difficulty of retrieval is likely core to what makes it effective in the first place. The fluency students experience when rereading material, as opposed to when they test themselves, can lead students to believe they have learned the material better than they really have, leading to over confidence. An indirect mechanism for why testing yields robust learning effects in classrooms is that it can help with students’ accurate assessments of what they do and do not know (ie, improve metacognition).¹⁵ For a more thorough consideration of possible psychological mechanisms for testing effect, refer to an article on this topic by Rowland.¹⁵

Implementing Tests as Learning Tools

When should educators choose multiple choice, fill-in-the-blank, or open-ended formats for examinations? Cognitively, each of these structures calls for a distinct form of memory retrieval. Multiple-choice questions require that students can recognize the correct response, fill-in-the-blank items require that students can recall

Table 1. Examples of Testing Effects Taken From Health Science Education

Aspect	Example Evidence
Test vs restudy	Physicians’ knowledge was repeatedly tested (versus restudy) and showed greater retention of material 6 months after the initial session ($d=.91$). ⁶⁸
	Dental students’ ability was tested (vs restudy) and showed greater diagnostic accuracy 1 week later after the initial session ($d=.33$). ⁶⁹
	Medical students were tested (vs restudy) and showed greater cardiac resuscitation skills ($d=.77$). ^{17,18}
Spacing of practice vs massed practice	Surgeons performed spaced practice (vs massed) performed better 1 year later on surgical procedures ($d=1.0$). ⁷²
	Student pharmacists performed spaced practice (vs massed) on learning brand/generic names and 6 weeks after the semester showed better retention ($d=.44$). ⁷³

information when given a cue or reminder (cued recall), and open-ended questions require students to generate a response with little or no provided structure (free recall). However, any format that encourages memory retrieval can improve learning better than restudying the information does (recognition, $d=0.36$; cued recall, $d=0.72$; free recall, $d=0.81$).^{15,16} Some researchers have argued that question formats requiring more effort to generate the answer on the part of the student are better for learning. Free recall tests can also help the learner practice developing mental pathways leading to the information they need to recall.²¹⁻²⁴ However, subsequent research has shown that multiple-choice questions can be equally effective.^{25,26} As instructors know, there is great variation within these categories of tests, and it is certainly possible to assign multiple-choice questions that require more effortful retrieval than fill-in-the-blank items. To recap, any format is helpful, and the specific structure chosen should be based on instructor needs and classroom context.

Within health science education, testing that encourages learning can take many forms beyond written assessments, including audience response systems (ARS) or “clickers,” questioning techniques, cooperative learning, and more traditional classroom assessment. Empirical studies of ARSs have shown positive effects in maintaining students’ attention but minimal impact on learning.²⁷ However, this may have more to do with how ARS questions are implemented in the classroom than a limitation of the system itself. Clicker questions are often used to assess content that was covered minutes before, which may be too soon to elicit a benefit of testing. Questions posed using ARSs will likely be more effective if they are implemented after a long delay. For example, clicker questions can be utilized at the end of class as review of material covered throughout that class period²⁸ or as a way to begin the following class.

The classic model of instructor-led questioning to a class can also be opportunity for retrieval practice. However, this approach yields a benefit only for the students who retrieve the information in response to the question. If, for example, a question is posed and an instructor calls on the first person who raises their hand to answer, many students in the class will not have had time to generate a response, functionally rendering the questioning more of a “study” event than a “test” event. Thus, one simple but powerful suggestion for implementing whole-class questioning as opportunities to generate a testing effect include waiting a sufficient amount of time for a response (at least 3 to 5 seconds^{29,30}) to increase the chance that other students will have time to practice retrieval. Of course, even with extended wait time, not all students will engage in retrieval when a question is posed

to a large class. Other approaches include having students use a classroom assessment technique (eg, PUREMEM³¹) or cooperative learning approach (eg, think-pair-share) that requires an action on the learner’s part. One benefit of think-pair-shares is that rather than one or two students retrieving and sharing their thinking, many students can practice retrieving and elaborating their answer out loud. This may be particularly helpful for learners who, while motivated to learn, may not be comfortable answering a question in front of the class but are willing to do so with a peer.

The testing effect also can be accomplished through simulation. As an example, when learners are using mannequins or standardized patients to emulate real-life clinical scenarios, there are often few explicit cues for memory, but there are the implicit cues of the patient’s signs and symptoms. In one research study, students first learned to diagnose and treat patients with three neurological conditions.³² Students then participated in one of three activities: interacting with a standardized patient, completing a written short-answer test, or studying a review sheet. About six months after initial learning, students completed two final tests administered one week apart on the three neurological conditions, which consisted of interacting with standardized patients and completing a written short-answer test. Two results proved notable. One was that students who interacted with the standardized patient as a learning activity performed better on the final tests with patients compared to students in the other two conditions (test or restudy). The second was that when examining performance on the final written test, students who had completed either the patient or written test learning activity conditions performed equally well and better than students who had studied a review sheet. This suggests that retrieval, even if mismatched with the final assessment modality, is better than re-study. Taken together, there are many avenues in which the testing effect can be implemented.

Another consideration is whether the benefits of testing help in transfer. That is, in order for a practice test to improve performance, does it need to be the same format as the final test?³³⁻³⁶ This could be relevant, for example, if a course uses short-answer questions but the licensure examination (eg, NAPLEX) uses multiple-choice items. One reason that practice tests that are matched in format to a subsequent test will do more to boost performance, an effect which might override the retrieval benefits of testing in various forms. Fortunately, research has shown that the benefit of repeated testing persists even when the format of the practice test does not match that of the final criterion test.^{36,37} Overall, there is little difference in outcomes when there is a mismatch between initial and final test (mismatch, $d=0.68$; matched, $d=0.64$).^{15,16}

Students in the health sciences take tests not to become experts at answering test questions. They do so to prepare to serve as excellent health professionals. This requires transferring their skills and knowledge from the original context of learning to another situation. Thus, another question related to the issue of transfer is if testing allows for transfer of knowledge. In other words, does the testing effect allow the learner to use that information outside of the context in which it was learned? Evidence suggests that testing does help with transfer. One laboratory study examined whether testing would help people transfer information about bat echolocation to a task about how submarines use sonar, which they were able to do successfully.³⁸ In the study noted previously that involved testing healthcare students using a pen and paper versus standardized patients,³² students demonstrated transfer in both directions (paper to patient and patient to paper). However, the effects were more robust when students saw the patient than when they took the written examination. More work is needed in this area within health science education, especially as it concerns the relationship between foundational science material and more clinical applications.

Related to the question of format is the question of how difficult tests should be. Because leading theories of the testing effect suggest that its effectiveness is related to difficulty, more difficult retrieval tasks will typically lead to better retrieval.³⁹⁻⁴¹ On the other hand, only successful retrieval attempts (or unsuccessful attempts with effective feedback) are likely to result in increased learning. This is to some extent intuitive, ie, a test on which students do not know any of the answers will not yield any benefit to learning. Thus, more difficult examinations will likely yield more lasting learning outcomes, to the extent that they can yield opportunities for successful retrieval.

Another aspect in enhancing the testing effect is feedback. Ideal feedback should be honest, specific, and timely, with direction on how to get better (testing with no feedback, $d=0.30-0.60$; testing with feedback, $d=0.60-0.73$).^{15,16} Feedback can improve the results of the testing effect and, in some cases, delayed feedback is superior to immediate feedback.⁴²⁻⁴⁵ One reason why delayed feedback may improve performance is that, through delay, the learner must retrieve that original thought process and that retrieval can strengthen memory. In other words, when students have to think about “why” they chose the answer they did, retrospection is it itself a retrieval attempt. However, if feedback is delayed too long, too much forgetting may occur or the motivation for received feedback may dissipate. Delayed feedback can take a variety of forms, such as responding to the muddiest points (asking students to write down what was most unclear or

most confusing during the class session), conducting post-examination reviews, and delaying feedback for online quizzes or assessments by a day.

While tests in a variety of formats result in improved performance relative to restudy, having multiple tests spaced over time has consistently been shown to have an advantage over a single test (massed testing).⁴⁶⁻⁴⁸ A single retrieval opportunity is better than none, but multiple retrievals, especially in a variety of contexts, produces greater long-term retention. When comparing practice that is spaced over time (eg, two sessions of five problems, one week apart) to massed practice (eg, 10 problems at once), the spaced condition leads to longer-term retention, although short-term performance measures may be similar or higher for the massed practice.^{46,49-51} For longer-term retention of information, research has shown the longer intervals between sessions tend to be more effective than shorter intervals.^{46,51,52} Other evidence suggests that the spacing interval should be 10% to 20% of the desired retention interval. That is, if an instructor wants learners to remember material or a skill for a year, the practice interval should be 1.2 to 2.4 months.^{46,53,54} However, it is not the specific titration of the retention interval that is most relevant. The larger principle, which has amassed much evidence in the psychological literature, is that spacing out practice results in better retention of information compared to massed practice. Thus, doing a little bit of retrieval each day is better than concentrating it in one day or one class session. This is important not only for students considering study strategies, but also for course and curriculum designers. As much as possible, instructors should design learning experiences cumulatively so that tests include prior material, not just the information discussed since the last test or course.

Repeated retrieval fits with the concept of *deliberate practice*, which is that deliberate effort to improve performance in a specific domain is critical to becoming an expert.^{55,56} Deliberate practice includes well-defined learning objectives that lead to repeated practice and clear outcome measures, ie, instructional alignment. This type of practice forms an iterative process of feedback and monitoring (metacognition) that leads to mastery. In the Best Evidence Medical Education (BEME) review of simulation-based education,⁵⁷ deliberate practice was found to be a key element leading to improvement in patient care. In a meta-analysis of simulation-based medical education, the authors demonstrated a combined effect in favor of improved skill learning through deliberate practice using simulation compared to traditional curricula.⁵⁸ Despite the clear benefits of deliberate practice, it is rarely applied in pharmacy curriculum. Incorporating tests as learning tools is one way to remedy this.

Thus far, we have discussed testing in general terms. Now we want to focus specifically on the potential value of multiple-choice questions (MCQs), which is a topic of debate among faculty members. Much research on MCQs focuses on how to appropriately construct multiple-choice tests to measure student learning and assess higher-order processing.⁵⁹ However, there is value in using multiple-choice questions as learning tools. The traditional view of multiple-choice questions is that they require less effort and simply require recognition of the correct answer, and thus offering fewer benefits to the learner. However, as noted above, the benefits of one format over another are not cut and dry, with some studies showing better benefits from using short-answer questions,²¹⁻²³ and others showing better results with using multiple-choice questions.^{25,26} However, there may be one benefit that multiple-choice questions have that other types of questions do not as we illustrate in the following hypothetical scenario: Imagine a student, Cristina, is trying to answer a short-answer question. Cristina does not know the answer. As a result, she will get that question incorrect and little learning will occur during this process. If Cristina had been given an MCQ, she might have been able to reason her way to choosing the correct response by retrieving information about the potential answer choices from memory and eliminating them as incorrect. Thus, even though she was not able to generate the correct answer on her own, she would still be learning because she would have to retrieve information about all the choices to eliminate them as options.

In a second scenario, imagine a student, Michelle, is also given a short-answer question. She knows the correct answer but is unable to call it to mind. During the examination, she experiences a “tip-of-the-tongue” feeling while attempting to retrieve the correct answer, which is “bilirubin.” Her thoughts might be, “The word starts with a ‘b’ and sounds like a child’s name.” If the question had been presented as an MCQ, when Michelle saw the answer choices, she would have recognized the word she was unable to recall (ie, bilirubin), eliminating the tip-of-the-tongue state she experienced when trying to respond to the short-answer question. The MCQ process would have helped stabilize that memory and would have resulted in her ability to generate that response later.⁹ Multiple-choice questions are a powerful way to stabilize access to knowledge that may have been learned well but cannot be retrieved when needed (marginal knowledge).^{9,60} This technique capitalizes on students’ ability to select the correct choice, re-exposing them to information, which has been used to stabilize prior knowledge in student-pharmacists from their prerequisite courses.⁶⁰ Thus, an MCQ can help remind students of material they once learned but

can no longer access easily because of disuse, such as material from prior course work.

From a cognitive science perspective, there are potential negative consequences to using MCQs. One issue is that the learner may remember an incorrect lure instead of the correct answer, using it to answer other related questions in the future.^{44,61,62} Note that this effect is seen when students explicitly choose the incorrect lure as the answer; exposure to the incorrect answers is not sufficient for producing this “negative testing effect.” Moreover, this potential consequence of MCQs is overshadowed by the positive aspects of the testing effect,^{43,44,61-63} and may be of more concern when a student has minimal baseline knowledge.^{35,61,64} A straightforward way in which this effect can be minimized or negated is by providing instructor feedback about what the correct choice is and why.^{62,65}

The beneficial effects of testing can also diminish over time.⁴⁴ However, researchers are investigating ways to improve retention. One approach involves designing MCQs with plausible lures requiring students to consider why each answer is correct and why each lure is incorrect.²⁵ One study demonstrated this approach had a greater positive influence on learning than free recall questions. This suggests that a good MCQ is one which requires more extensive retrieval rather than recognition of the correct answer. While there is no “roadmap” to constructing these types of questions, Table 2 offers some general guidance from the literature.

DISCUSSION

We split our list of concrete recommendations into two bins to reflect ideas for instructors in designing course and curricula (Table 3), as well as strategy approaches that can be recommended for students (Table 4). None of these are necessarily “must-do” items. Rather, we consider them collectively as a toolbox from which instructors and students can select their favorite tools to hone over time.

Our goal was to provide a backdrop of psychological theory in which to ground the recommendations here related to testing for learning, and especially related to multiple-choice testing. Given the diversity of students, curricula, educators, and context within any medical education curriculum, we believe imparting a backdrop of information is much more likely to yield practical gains than listing essential recommendations. In that spirit, we offer an illustrative example of what the principles described above might look like on the first few days of a course:

Classroom Vignette: On the first day of class, students complete a multiple-choice question assessment related to prior knowledge, which students will need for Dr. Fuller’s

Table 2. Tips for Making the Most of Multiple-Choice Questions Administered to Doctor of Pharmacy Students

Tip	Rational and Evidence
Create lures that are reasonable alternatives	Plausible lures are more likely to yield more active retrieval processes, rather than recognition of the correct answer. ^{62,74}
Use 2-3 lures in addition to the correct answer	Fewer lures means learners are less likely to endorse a lure which minimizes the negative testing effect. ^{62,74} More lures (>2) do not add to better psychometric properties. ⁷⁵⁻⁸⁰ It also is quicker to proceed through an examination with fewer lures. ⁷⁷
Avoid “none of the above” and “all of the above” as answer choices	While more lures might present a desirable difficulty for learning, ⁶⁵ this may require the lures be plausible. When “none-of the above” is the correct choice, there is no testing benefit because the correct answer was not present. In fact, students are more likely to use the incorrect lures later as correct choice. ⁸¹
Avoid complex format multiple-choice questions (eg, K-type questions)	These types of questions inadvertently enable test-takers to engage in strategic guessing leading to higher performance with lower reliability. ^{82,83} In addition, these items are not better at measuring higher-order thinking than simpler item types. ⁸⁴ Overall, the costs associated with complex formats greatly outweigh the benefits. ^{84,85}
Following guidelines for constructing multiple-choice tests	Use National Board of Medical Examiners guidelines to constructing multiple-choice questions. ⁸⁵ This helps thwart learners who may choose the correct answer from being test-savvy (eg, avoiding answers that are not grammatically aligned with the question stem), rather than successfully retrieving the information.

course. Dr. Fuller includes this to assess what students remember from previous courses and to reactivate any marginal knowledge, ie, information which is stored in memory but may have become inaccessible with time. After students complete the assessment, she verbally connects this old information with how it will be used and applied in the current course. Finally, she instructs students on the material they should learn for the next class period.

At the start of the next class period, students complete a short multiple-choice quiz on the material they

were asked to learn. Dr. Fuller does this to keep students accountable for their learning and to use the quiz as a retrieval practice to help consolidate this information into memory. She uses a variety of activities during the remainder of class, including think-pair-shares and a case vignette followed by a Socratic discussion, and ends with a summary of the learning objectives and feedback on what the class seems to be grasping well and where they may need to devote more attention. She chose these activities to help students retrieve the information they

Table 3. Ideas for Instructors and Course Designers to Incorporate Testing as a Learning Tool in the Doctor of Pharmacy Curriculum

Utilize frequent low-stakes assessments. For example, consider beginning lessons or class periods with a multiple-choice test to reactivate less retrievable knowledge. ⁹
Vary the recall format of questions (recognition, cued recall, free recall) by incorporating a mix of multiple-choice, fill-in-the-blank, and/or short-answer item. ⁸⁶
Use clickers to review information at a delay, such as information from the previous class period, rather than information that was recently presented as a check for attention.
Carry through important points from prior classes by testing cumulatively, which is one way to implement spacing out testing over time. ^{50,87}
Provide feedback on questions and reasoning behind why incorrect choices (ie, lures) are incorrect using strategies such as examination reviews. ⁸⁸ Consider delaying feedback after testing.
Provide practice tests for students to do on their own.

Table 4. Strategies for Doctor of Pharmacy Students to Incorporate More Retrieval Practice Into Study and Assessment to Foster Longer-Term Retention

Use instructor-provided practice examinations, if available.
Convert learning objectives to questions and generate responses.
Write a few quiz questions pertaining to each class section or textbook chapter. Collaborate with classmates to generate a practice test to test each other.
Practice free recall of class content without looking at notes, then use notes to fill in what you missed. ⁸⁹
Use flashcards to alternate between study (review of material) and testing (retrieval) over time until concepts are learned. Do not drop successfully retrieved material immediately and consider repeating everything over time. ¹⁴
Do not self-test on material that was just studied – wait at least an hour.
When completing multiple-choice questions, write explanations for why each choice is correct or incorrect.
When taking a test, write down your confidence in the answer. Revisit the ones with low confidence. ⁹⁰

learned prior to class, and to have them work with the information in a different context. Throughout the activities, she provides feedback.

The third class period begins with Dr. Fuller presenting three “clicker” questions covering the information discussed in the last class period to help students retrieve information from that session, all the while knowing that students have forgotten large portions of the content. Based on the results, she then does a five-minute review on information presented the previous week and then continues with a case that is slightly different than the prior class. For this case, students work individually and then pair up with their classmates for discussion. Dr. Fuller uses the case to have students retrieve prior knowledge and apply it in a different context. After allowing students time to construct their answers, she debriefs the class to explain their answers and offers corrective feedback. After the debrief, she has students complete a muddiest point exercise. She does this individually to ensure all students are engaged, and then she has the students meet in small groups for discussion and feedback. She debriefs the class to better understand what students understand and uses the muddiest point exercise once again to help students retrieve information and reflect on what they still feel is unclear; this process of reflecting and giving the instructor feedback also serves as an aid for students’ metacognition.

Throughout the remainder of the course Dr. Fuller uses additional formative assessments to provide students feedback on their learning and provides feedback on summative assessments. She invests considerable effort in helping students retrieve information from memory from a variety of perspectives and contexts.

CONCLUSION

Testing can have multiple learning benefits. The format of tests and their timing requires thoughtfulness on the instructor’s part, ie, to ask the right questions, in the right format, at the right time, that is equivalent to the

personalized medicine mantra of the right dose at the right time for the right person. Our main point was to emphasize that practice recalling knowledge and skills is at least as important as how students are initially exposed to information. Moreover, learning is an iterative process. After an expertly delivered lecture or an intense study session, even the most well-learned information requires regular retrieval for long-term retention. The study strategies that feel the most comfortable for students are precisely those which encourage short-term rather than long-term retention. Finally, given the constraints of a study session or classroom, testing as a learning strategy is a cost-effective way to spend the precious resource of time, as implementing a more cognitively demanding strategy such as testing outweighs a more comfortable strategy such as re-study. To summarize, we stress that incorporating opportunities for retrieval after teaching is an essential component of lasting learning. We hope that the recommendations offered here provide a fertile starting place for incorporating testing into each stage of the learning process.

REFERENCES

1. Landauer TK. How much do people remember? *Cogn Sci*. 1986;10:477-493.
2. Shiffrin RM, Atkinson RC. Storage and retrieval processes in long-term memory. *Psychol Rev*. 1969;76(2):179-193. doi:10.1037/h0027277.
3. Tulving E. Cue-dependent forgetting: when we forget something we once knew, it does not necessarily mean that the memory trace has been lost; it may only be inaccessible. *Am Sci*. 1974;62(1):74-82.
4. Born J, Wilhelm I. System consolidation of memory during sleep. *Psychol Res*. 2012;76(2):192-203. doi:10.1007/s00426-011-0335-6.
5. Clopath C. Synaptic consolidation: an approach to long-term learning. *Cogn Neurodyn*. 2012;6(3):251-257. doi:10.1007/s11571-011-9177-6.
6. De Schrijver S, Barrouillet P. Consolidation and restoration of memory traces in working memory. *Psychon Bull Rev*. 2017;24(5):1651-1657. doi:10.3758/s13423-017-1226-7.

7. Nadel L, Hubbach A, Gomez R, Newman-Smith K. Memory formation, consolidation and transformation. *Neurosci Biobehav Rev.* 2012;36(7):1640-1645. doi:10.1016/j.neubiorev.2012.03.001.
8. Winocur G, Moscovitch M. Memory transformation and systems consolidation. *J Int Neuropsychol Soc.* 2011;17(5):766-780. doi:10.1017/s1355617711000683.
9. Cantor AD, Eslick AN, Marsh EJ, Bjork RA, Bjork EL. Multiple-choice tests stabilize access to marginal knowledge. *Mem Cognit.* 2015;43(2):193-205. doi:10.3758/s13421-014-0462-6.
10. Tulving E, Pearlstone Z. Availability versus accessibility of information in memory for words. *J Verbal Learn Verbal Behav.* 1996;5:381-391.
11. Brown AS. A review of the tip-of-the-tongue experience. *Psychol Bull.* 1991;109(2):204-223.
12. Bjork RA, Bjork EL. A new theory of disuse and an old theory of stimulus fluctuation. In: Estes WK, Healy AF, Kosslyn SM, Shiffrin RM, eds. *Essays in Honor of William K. Estes.* Hillsdale, NJ: L. Erlbaum Associates; 1992. <http://www.loc.gov/catdir/enhancements/fy0742/91039697-d.html>.
13. Brown PC. *Make It Stick: The Science of Successful Learning.* Cambridge, MA: The Belknap Press Harvard University Press; 2014.
14. Miyatsu T, Nguyen K, McDaniel MA. Five popular study strategies: their pitfalls and optimal implementations. *Perspect Psychol Sci.* 2018;13(3):390-407.
15. Rowland CA. The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychol Bull.* 2014;140(6):1432.
16. Adesope OO, Trevisan DA, Sundararajan N. Rethinking the use of tests: a meta-analysis of practice testing. *Rev Educ Res.* 2017;87(3):659-701. doi:10.3102/0034654316689306.
17. Bjork RA. Memory and metamemory considerations in the training of human beings. In: Metcalfe J, Shimamura A, eds. *Metacognition: Knowing about Knowing.* Cambridge, MA: MIT Press; 1994:185-205.
18. Roediger HL, Karpicke JD. Reflections on the resurgence of interest in the testing effect. *Perspect Psychol Sci.* 2018;13(2):236-241.
19. Karpicke JD, Roediger HL. The critical importance of retrieval for learning. *Science.* 2008;319(5865):966-968. doi:10.1126/science.1152408.
20. Roediger HL, Karpicke JD. Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci.* 2006;17(3):249-255. doi:10.1111/j.1467-9280.2006.01693.x.
21. Butler AC, Roediger HL. Testing improves long-term retention in a simulated classroom setting. *Eur J Cogn Psychol.* 2007;19(4-5):514-527. doi:10.1080/09541440701326097.
22. Kang SHK, McDermott KB, Roediger HL. Test format and corrective feedback modify the effect of testing on long-term retention. *Eur J Cogn Psychol.* 2007;19(4-5):528-558. doi:10.1080/09541440601056620.
23. McDaniel MA, Anderson JL, Derbish MH, Morrisette N. Testing the testing effect in the classroom. *Eur J Cogn Psychol.* 2007;19(4-5):494-513. doi:10.1080/09541440701326154.
24. Glover JA. The "testing" phenomenon: Not gone but nearly forgotten. *J Educ Psychol.* 1989;81(3):392-399. doi:10.1037//0022-0663.81.3.392.
25. Little JL, Bjork EL, Bjork RA, Angello G. Multiple-choice tests exonerated, at least of some charges: fostering test-induced learning and avoiding test-induced forgetting. *Psychol Sci.* 2012;23(11):1337-1344. doi:10.1177/0956797612443370.
26. McDermott KB, Agarwal PK, D'Antonio L, Roediger HL, McDaniel MA. Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *J Exp Psychol-Appl.* 2014;20(1):3-21. doi:10.1037/xap0000004.
27. Hunsu NJ, Adesope O, Bayly DJ. A meta-analysis of the effects of audience response systems (clicker-based technologies) on cognition and affect. *Comput Educ.* 2016;94:102-119. doi:10.1016/j.compedu.2015.11.013.
28. Davis SD, Chan JCK, Wilford MM. The dark side of interpolated testing: frequent switching between retrieval and encoding impairs new learning. *J Appl Res Mem Cogn.* 2017;6(4):434-441.
29. Barrett M, Magas CP, Gruppen LD, Dedhia PH, Sandhu G. It's worth the wait: optimizing questioning methods for effective intraoperative teaching. *ANZ J Surg.* 2017;87(7-8):541-546. doi:10.1111/ans.14046.
30. Cho. YH, Lee SY, Jeong DW, et al. Analysis of questioning technique during classes in medical education. *BMC Med Educ.* 2012;12(1):39-39. doi:10.1186/1472-6920-12-39.
31. Lyle KB, Crawford NA. Retrieving essential material at the end of lectures improves performance on statistics exams. *Teach Psych.* 2011;38(2):94-97. doi:10.1177/0098628311401587.32.
32. Larsen DP, Butler AC, Lawson AL, Roediger Iii HL. The importance of seeing the patient: test-enhanced learning with standardized patients and written tests improves clinical application of knowledge. *Adv Health Sci Educ.* 2013;18(3):409-425. doi:10.1007/s10459-012-9379-7.
33. Carpenter SK, Delosh EL. Impoverished cue support enhances subsequent retention: support for the elaborative retrieval explanation of the testing effect. *Mem Cognit.* 2006;34(2):268-276. doi:10.3758/bf03193405.
34. Hinze SR, Wiley J. Testing the limits of testing effects using completion tests. *Memory.* 2011;19(3):290-304. doi:10.1080/09658211.2011.560121.
35. Marsh EJ, Agarwal PK, Roediger HL. Memorial consequences of answering SAT II questions. *J Exp Psychol Appl.* 2009;15(1):1-11. doi:10.1037/a0014721.
36. McDaniel MA, Roediger HL, McDermott KB. Generalizing test-enhanced learning from the laboratory to the classroom. *Psychon Bull Rev.* 2007;14(2):200-206. doi:10.3758/bf03194052.
37. McConnell MM, St-Onge C, Young ME. The benefits of testing for learning on later performance. *Adv Health Sci Educ.* 2015;20(2):305-320. doi:10.1007/s10459-014-9529-1
38. Butler AC. Repeated testing produces superior transfer of learning relative to repeated studying. *J Exp Psychol Learn Mem Cogn.* 2010;36(5):1118-1133. doi:10.1037/a0019902.
39. Brown R, McNeill D. The "tip of the tongue" phenomenon. *J Verbal Learn Verbal Behav.* 1966;5(4):325-337.
40. Craik FIM. The fate of primary memory items in free recall. *J Verbal Learn Verbal Behav.* 1970;9(2):143-148. doi:10.1016/S0022-5371(70)80042-1.
41. Gardiner FM, Craik FIM, Bleasdale FA. Retrieval difficulty and subsequent recall. *Mem Cognit.* 1973;1(3):213-216. doi:10.3758/BF03198098.
42. Butler AC, Karpicke JD, Roediger HL. The effect of type and timing of feedback on learning from multiple-choice tests. *J Exp Psychol Appl.* 2007;13(4):273-281. doi:10.1037/1076-898x.13.4.273.
43. Butler AC, Roediger HL. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Mem Cognit.* 2008;36(3):604-616. doi:10.3758/mc.36.3.604.
44. Fazio LK, Agarwal PK, Marsh EJ, Roediger HL. Memorial consequences of multiple-choice testing on immediate and delayed tests. *Mem Cognit.* 2010;38(4):407-418. doi:10.3758/MC.38.4.407.
45. Mullet HG, Butler AC, Verdin B, von Borries R, Marsh EJ. Delaying feedback promotes transfer of knowledge despite student

- preferences to receive feedback immediately. *J Appl Res Mem Cogn*. 2014;3(3):222-229. doi:10.1016/j.jarmac.2014.05.001.
46. Cepeda NJ, Pashler H, Vul E, Wixted JT, Rohrer D. Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychol Bull*. 2006;132(3):354-380. doi:10.1037/0033-2909.132.3.354.
47. Rawson KA, Dunlosky J. Optimizing schedules of retrieval practice for durable and efficient learning: how much is enough? *J Exp Psychol Gen*. 2011;140(3):283-302. doi:10.1037/a0023956.
48. Rawson KA, Dunlosky J. When is practice testing most effective for improving the durability and efficiency of student learning? *Educ Psychol Rev*. 2012;24(3):419-435. doi:10.1007/s10648-012-9203-1.
49. Benjamin AS, Tullis J. What makes distributed practice effective? *Cognit Psychol*. 2010;61(3):228-247. doi:10.1016/j.cogpsych.2010.05.004.
50. Carpenter SK, Cepeda NJ, Rohrer D, Sean HKK, Pashler H. Using spacing to enhance diverse forms of learning: review of recent research and implications for instruction. *Educ Psychol Rev*. 2012;24(3):369-378. doi:10.1007/s10648-012-9205-z.
51. Cepeda NJ, Coburn N, Rohrer D, Wixted JT, Mozer MC, Pashler H. Optimizing distributed practice theoretical analysis and practical implications. *Exp Psychol*. 2009;56(4):236-246. doi:10.1027/1618-3169.56.4.236.
52. Karpicke JD, Bauernschmidt A. Spaced retrieval: absolute spacing enhances learning regardless of relative spacing. *J Exp Psychol Learn Mem Cogn*. 2011;37(5):1250-1257. doi:10.1037/a0023436.
53. Küpper-Tetzel CE, Kapler IV, Wiseheart M. Contracting, equal, and expanding learning schedules: the optimal distribution of learning sessions depends on retention interval. *Mem Cognit*. 2014;42(5):729-741. doi:10.3758/s13421-014-0394-1.
54. Pyc MA, Rawson KA. Testing the retrieval effort hypothesis: does greater difficulty correctly recalling information lead to higher levels of memory? *J Mem Lang*. 2009;60(4):437-447. doi:10.1016/j.jml.2009.01.004.
55. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev*. 1993;100(3):363-406. doi:10.1037/0033-295x.100.3.363.
56. Macnamara BN, Hambrick DZ, Oswald FL. Deliberate practice and performance in music, games, sports, education, and professions: a meta-analysis. *Psychol Sci*. 2014;25(8):1608-1618. doi:10.1177/0956797614535810.
57. Issenberg BS, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach*. 2005;27(1):10-28. doi:10.1080/01421590500046924.
58. McGaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? a meta-analytic comparative review of the evidence. *Acad Med*. 2011;86(6):706-711. doi:10.1097/ACM.0b013e318217e119.
59. Frederiksen N. The real test bias: influences of testing on teaching and learning. *Am Psychol*. 1984;39(3):193-202. doi:10.1037/0003-066x.39.3.193.
60. Butler AC, Cantor AD, Campbell K, Marsh EJ, Persky AM. Stabilizing access to marginal knowledge in a classroom setting. *Appl Cogn Psychol*. Under Review.
61. Marsh EJ, Roediger HL, Bjork RA, Bjork EL. The memorial consequences of multiple-choice testing. *Psychon Bull Rev*. 2007;14(2):194-199. doi:10.3758/bf03194051.
62. Roediger HL, Marsh EJ. The positive and negative consequences of multiple-choice testing. *J Exp Psychol Learn Mem Cogn*. 2005;31(5):1155-1159. doi:10.1037/0278-7393.31.5.1155.
63. Fazio LK, Agarwal PK, Marsh EJ, Roediger HL. Memorial consequences of multiple-choice testing on immediate and delayed tests. *Mem Cognit*. 2010;38(4):407-418. doi:10.3758/mc.38.4.407.
64. Marsh EJ, Fazio LK, Goswick AE. Memorial consequences of testing school-aged children. *Memory*. 2012;20(8):899-906. doi:10.1080/09658211.2012.708757.
65. Butler AC, Marsh EJ, Goode MK, Roediger HL. When additional multiple-choice lures aid versus hinder later memory. *Appl Cogn Psychol*. 2006;20(7):941-956. doi:10.1002/acp.1239.
66. Goldstein EB. *Cognitive Psychology: Connecting Mind, Research, and Everyday Experience*. 5th ed. Boston, MA: Cengage Learning; 2018.
67. Norris D. Short-term memory and long-term memory are still different. *Psychol Bull*. 2017;143(9):992-1009. doi:10.1037/bul000108.
68. Larsen DP, Butler AC, Roediger Iii HL. Repeated testing improves long-term retention relative to repeated study: a randomized controlled trial. *Med Educ*. 2009;43(12):1174-1181. doi:10.1111/j.1365-2923.2009.03518.x.
69. Baghdady M, Carnahan H, Lam EWN, Woods NN. Test-enhanced learning and its effect on comprehension and diagnostic accuracy. *Med Educ*. 2014;48(2):181-188. doi:10.1111/medu.12302.
70. Kromann CB, Bohnstedt C, Jensen ML, Ringsted C. The testing effect on skills learning might last 6 months. *Adv Health Sci Educ*. 2010;15(3):395-401. doi:10.1007/s10459-009-9207-x.
71. Kromann CB, Jensen ML, Ringsted C. The effect of testing on skills learning. *Med Educ*. 2009;43(1):21-27. doi:10.1111/j.1365-2923.2008.03245.x.
72. Spruit EN, Band GPH, Hamming JF. Increasing efficiency of surgical training: effects of spacing practice on skill acquisition and retention in laparoscopy training. *Surg Endosc*. 2015;29(8):2235-2243. doi:10.1007/s00464-014-3931-x.
73. Terenyi J, Anksorus HN, Persky AM. Impact of spacing of practice on learning brand name and generic drugs. *Am J Pharm Educ*. 2018;82(1):Article 6179. <https://doi.org/10.5688/ajpe6179>.
74. Marsh EJ, Cantor AD. Learning from the test: dos and don'ts for using multiple-choice tests. In: McDaniel MA, Frey RF, Fitzpatrick SM, Roediger Iii HL, eds. *Integrating Cognitive Science with Innovative Teaching in STEM Disciplines*. Washington UP; 2014. doi:10.7936/K7Z60KZK.
75. Gierl MJ, Bulut O, Guo Q, Zhang X. Developing, analyzing, and using distractors for multiple-choice tests in education: a comprehensive review. *Rev Educ Res*. 2017;87(6):1082-1116. doi:10.3102/0034654317726529.
76. Kilgour JM, Tayyaba S. An investigation into the optimal number of distractors in single-best answer exams. *Adv Health Sci Educ*. 2016;21(3):571-585. doi:10.1007/s10459-015-9652-7.
77. Schneid SD, Armour C, Park YS, Yudkowsky R, Bordage G. Reducing the number of options on multiple-choice questions: response time, psychometrics and standard setting. *Med Educ*. 2014;48(10):1020-1027. doi:10.1111/medu.12525.
78. Kilgour JM, Tayyaba S. An investigation into the optimal number of distractors in single-best answer exams. *Adv Health Sci Educ*. 2016;21(3):571-585. doi:10.1007/s10459-015-9652-7.
79. Schneid SD, Armour C, Park YS, Yudkowsky R, Bordage G. Reducing the number of options on multiple-choice questions: response time, psychometrics and standard setting. *Med Educ*. 2014;48(10):1020-1027. doi:10.1111/medu.12525.

80. Odegard TN, Koen JD. "None of the above" as a correct and incorrect alternative on a multiple-choice test: implications for the testing effect. *Memory*. 2007;15(8):873-885. doi:10.1080/09658210701746621.
81. Albanese MA. Type k and other complex multiple-choice items: an analysis of research and item properties. *Educ Meas Issues Pract*. 1993;12(1):28-33. doi:10.1111/j.1745-3992.1993.tb00521.x.
82. Albanese MA. Multiple-choice items with combinations of correct responses: a further look at the type k format. *Eval Health Prof*. 1982;5(2):218-228. doi:10.1177/016327878200500207.
83. Butler AC. Multiple-choice testing in education: are the best practices for assessment also good for learning? *J Appl Res Mem Cogn*. 2018. doi:10.1016/j.jarmac.2018.07.002. AU: please complete citation
84. Billings MS, DeRuchie K, Haist SA, et al. *Constructing Written Test Questions for the Basic and Clinical Sciences*. National Board of Medical Examiners; Philadelphia, PA. 1996, Updated 2016. https://health.uconn.edu/faculty-development/wp-content/uploads/sites/69/2017/06/constructing_written_test_questions.pdf. Accessed September 26, 2019.
85. Glass AL, Sinha N. Multiple-choice questioning is an efficient instructional methodology that may be widely implemented in academic courses to improve exam performance. *Curr Dir Psychol Sci*. 2013;22(6):471-477. doi:10.1177/0963721413495870.
86. Rawson KA, Dunlosky J. Relearning attenuates the benefits and costs of spacing. *J Exp Psychol*. 2013;142(4):1113.
87. Marsh EJ, Lozito JP, Umanath S, Bjork EL, Bjork RA. Using verification feedback to correct errors made on a multiple-choice test. *Memory*. 2012;20(6):645-653. doi:10.1080/09658211.2012.684882.
88. Putnam AL, Sungkhasettee VW, Roediger HL. Optimizing learning in college: tips from cognitive psychology. *Perspect Psychol Sci*. 2016;11(5):652-660. doi:10.1177/1745691616645770.
89. Couchman JJ, Miller NE, Zmuda SJ, Feather K, Schwartzmeyer T. The instinct fallacy: the metacognition of answering and revising during college exams. *Metacognition Learn*. 2016;11(2):171-185. doi:10.1007/s11409-015-9140-8.