



Published in final edited form as:

Anat Sci Educ. 2011 September ; 4(5): 243–248. doi:10.1002/ase.250.

Testing to enhance retention in human anatomy

Jessica M. Logan¹, Andrew J. Thompson^{1,2}, and David W. Marshak³

¹Department of Psychology, Rice University, Houston, Texas

²Baylor College of Medicine, Houston, Texas

³Department of Neurobiology and Anatomy, University of Texas Medical School, Houston, Texas

Abstract

Recent work in cognitive psychology has shown that repeatedly testing one's knowledge is a powerful learning aid and provides substantial benefits for retention of the material. To apply this in a human anatomy course for medical students, 39 fill-in-the-blank quizzes of about 50 questions each, one for each region of the body and four about the nervous system, were developed. The quizzes were optional, and no credit was awarded. They were posted online using Blackboard, which provided feedback, and they were very popular. To determine whether the quizzes had any effect on retention, they were given in a controlled setting to 21 future medical and dental students. The weekly quizzes included questions on regional anatomy and an expanded set of questions on the nervous system. Each question about the nervous system was given three times, in a slightly different form each time. The second quiz was given approximately half an hour after the first one, and the third was given one week after the second to assess retention. The quizzes were unpopular, but students showed robust improvement on the questions about the nervous system. The scores increased by almost 9% on the second quiz, with no intervention except viewing the correct answers. The scores were 29% higher on the third quiz than on the first, and there was also a positive correlation between the grades on the quizzes and the final examination. Thus, repeated testing is an effective strategy for learning and retaining information about human anatomy.

Keywords

anatomy education; human gross anatomy; assessment; E-learning; retention

INTRODUCTION

Instructors and students alike are very familiar with the use of testing in the classroom as a tool to assess mastery of knowledge and to assign grades in a course. However, testing can also be used as a powerful learning device that has substantial impacts on later retention (for a review, see Roediger et al., 2010). The benefits of repeated testing have been found with a

Correspondence to: Dr. David W. Marshak, Department of Neurobiology and Anatomy, University of Texas Medical School, PO Box 20708 Houston, TX 77225, USA. david.w.marshak@uth.tmc.edu, Phone: 713-500-5617, Fax: 713-500-0621.

JESSICA M. LOGAN, Ph.D., is an assistant professor of psychology in the Department of Psychology at Rice University in Houston, Texas. She teaches courses on research methods and memory, and she conducts research on the role of distributed practice and self-testing in learning, among other topics.

ANDREW J. THOMPSON, B.A., is presently a first-year medical student at Baylor College of Medicine in Houston, Texas. As an undergraduate student, he worked with Dr. Logan in the Department of Psychology at Rice University.

DAVID W. MARSHAK, Ph.D., is a professor of neurobiology and anatomy in the Department of Neurobiology and Anatomy, University of Texas Medical School in Houston, Texas. He teaches anatomy to first-year medical students and does research on neural circuits in the retina.

variety of materials, including word lists (Thompson et al., 1978), associated word pairs (Logan and Balota, 2008), foreign vocabulary words (Carrier and Pashler, 1992; Karpicke and Roediger, 2008), prose passages (Glover, 1989; Chan et al., 2006; Roediger and Karpicke, 2006a), and lectures (Butler and Roediger, 2008). One influential study by Karpicke and Roediger (2008) analyzed the effect of repeated testing versus repeated studying on the ability of students to remember the meanings of Swahili words. In their laboratory study, after a Swahili word was correctly defined on a test it was either: repeatedly studied and tested, repeatedly tested but not studied again (repeated testing), repeatedly studied but not tested again (repeated studying), or else neither studied nor tested again (selective studying). Initially, the learning curves for the four groups were identical; both repeated studying and repeated testing benefited memory performance. But when students were tested again after one week, the results were surprising. There was a large, positive benefit in retention of information from repeated testing, but not from studying. The latter two strategies involving repeated studying or selective studying, which are commonly used by students and essentially the norm in educational settings, were not effective in promoting retention of the material. Scores dropped to about 30% correct under both the repeated studying and the selective studying conditions. On the other hand, repeated testing produced a large positive effect; the results were essentially the same as with the repeated studying plus testing condition, about 80% correct. These findings highlight the ultimate benefits of testing, even without further studying. Another interesting finding was that the students participating in the experiment were not aware of the benefits of testing or the futility of repeated studying. In a survey given after learning the words initially, the students all predicted that they would remember the material equally well. Although this was a study conducted in the laboratory, there is also evidence that the benefits of testing can translate well into learning in the classroom (Butler and Roediger, 2007; McDaniel et al., 2007).

Several recent studies highlight the importance of augmenting traditional teaching methods in anatomy education to counteract the historical trend of allotting less time in human anatomy instruction in medical schools (Bergman et al. 2008, Drake et al., 2009; Sugand et al., 2010). One of the remedies suggested by Sugand and colleagues (2010) is to increase usage of the internet in teaching anatomy. Marsh et al. (2008) successfully used web-based learning modules to teach and quiz students about embryonic development, for example. Anatomy educators are also increasingly turning to research in the psychology of learning and memory to help guide innovations in medical school classrooms. There has been an emphasis on creating a more active learning environment for students where they more directly work with the information presented to them, rather than more traditional and passive lecture/study models (Bergman et al., 2008). Moreover, Bergman et al. (2008) make the point that revisiting topics already taught is important for retention of learning in anatomy classes. The current study was designed to implement many of these suggestions and to determine their effectiveness in anatomy education. The project incorporated the technique of repeated testing, so that topics related to the nervous system could be revisited and built upon over time. In addition, the quizzes were given in an online format to test the feasibility of augmenting classroom time with an instructional aid through the internet. Finally, the students were given immediate feedback about their quizzes and encouraged to seek assistance from an instructor.

In a pilot study, 12 practice quizzes with approximately 50 questions each were written for the material in the first one-third of the Human Anatomy course, Block 1, in 2009. The Human Anatomy course was team-taught by seven instructors to about 230 first-year medical students. In Block 1, there were 12 laboratories and 19 lectures, including clinical correlates; regions covered included the back, upper extremity, thorax and part of the neck. Block 1 had one written examination (50 questions), one practical examination (50 questions) and one peer teaching evaluation. Block 2 and Block 3 were organized similarly

and covered the remaining regions of the body. The quizzes were used only to promote learning and not for assessment; in other words, quizzes did not count toward final grades. The questions were straightforward and dealt only with basic anatomy. In order to maximize the effort required in answering the questions, they were in the “fill-in-the-blank” format, and they were presented one at a time, in random order using Blackboard content management system, version 8 (Blackboard, Inc., Washington, DC). After the quiz, feedback, including the students’ answers and the correct answers, was automatically provided.

After the Block 1 midterm examination, the faculty met with a group of student representatives, who reported that their classmates found the quizzes helpful and asked for more. First, similar quizzes were written to be used during Block 3, and then quizzes for Block 2 were written to aid the students in preparing for the final examination. By the end of the course, there was one quiz for each of the 35 regions of the body and four that dealt solely with the nervous system. No credit was offered for completing the practice quizzes, and the results were not monitored.

The outcomes of the optional practice quizzes for the 2009 first-year medical students were measured using a course evaluation with a Likert scale, and the practice quizzes were quite popular. Of the 230 medical students in the class, 92% of students tried them, and 79% of the students who used them found the practice quizzes helpful or somewhat helpful in learning the material. In an online evaluation of the course, there were no negative comments, and three students wrote anonymous, positive comments. For example, one student responded, “The quizzes were hard... but they really help to reinforce all of the concepts and to help you figure out your deficiencies.” Other students said much the same thing, including that the quizzes “helped me look at the material another way” and “they make studying Gross Anatomy fun and enjoyable.”

The first-year medical students took the practice quizzes under uncontrolled and, presumably, widely-varying conditions. For example, one student who was having academic difficulties mentioned during a counseling session that she had taken the quizzes with the book open, negating the benefit provided by effortful retrieval of the answers (Carpenter, 2009; Karpicke and Roediger, 2008). Thus, it was impossible to measure the effects of the practice quizzes on learning. A rigorous analysis would require carefully-monitored testing conditions, but there is no facility with secure computers that is large enough to accommodate all of the first-year medical students at the University of Texas Medical School-Houston.

Thus, a follow-up study was conducted in a more controlled setting in order to assess the potential impact of these quizzes on learning in an academic environment. In this second study, data were collected from a smaller group of students, the members of the Pre-entry Program, designed to help students accepted into medical or dental school prepare for the first year. There were two goals for these experiments: (1) to determine whether repeated testing was an effective strategy for learning and remembering anatomy and (2) to determine whether any benefits of testing would persist over time.

Method

Participants—In July 2010, 21 incoming first-year medical and dental students took part in the Pre-entry Program, which included lectures and laboratories in human anatomy. The majority of the students in the Pre-entry Program were planning to enroll in the University of Texas Medical School at Houston in a few weeks; there were eight men and eight women (nine White, two Hispanic, one Asian, three Black and one Native American). Their mean grade point average was 3.68 (range: 3.5-4.0), and their mean MCAT score was 29.9 (range:

24-37). There were five future University of Texas dental students in the study, four men and one woman; all were Hispanic. Their mean grade point average was 3.67 (range: 3.35-3.96), and their mean DAT score was 19.2 (range: 16-22). Some of the accepted first-year students had been invited to participate in the program based on several criteria suggesting that they might be at risk of academic difficulty, but because the program was open to all first-year students, some participants were not considered to be at risk.

Because the practice quizzes on Blackboard and the course evaluation are used routinely, The University of Texas Health Science Center Committee for the Protection of Human Subjects determined that this research was exempt from Institutional Review Board review. To ensure that the study was ethical, at the first class meeting, students were told about the practice quizzes and this research study. It was explained that this was a research study being conducted during class, and the students had the option of not participating in the study. The students were shown a “letter of information” via Blackboard and told that they could opt out of the study during any of the three sessions. All 21 students responded positively to the letter of information and agreed to participate in the study.

Materials and Procedure—The material taught was covered by six quizzes of about 50 questions each. Three Friday afternoon sessions were conducted in a classroom equipped with 30 computers arranged so that the screens were visible to the teaching assistants, who served as proctors. Each quiz lasted approximately half an hour, and two quizzes were given in a session. In addition, an instructor was there to help with the content, and experts from the Office of Educational Programs were there to help with Blackboard issues. The regional anatomy questions were unchanged from the pilot study, but there was an expanded set of nervous system questions that would serve as the targets for repeated testing. Rather than asking the exact same question three times, three versions of each nervous system question were created. The versions of the questions were counterbalanced. Students were randomly assigned to one of six groups, which varied in the order in which the three versions were presented and also varied in whether odd-numbered or even-numbered questions were repeated first. An example is provided below, with correct answers given in parentheses. However, the answers were not shown to students until after they completed the entire quiz.

- ▲ Quiz 1: “Preganglionic cell bodies in the parasympathetic system are found at the ____ level of the spinal cord.” (sacral)
- ▲ Quiz 2: “The parasympathetic neurons at the sacral level of the spinal cord are ____-ganglionic.” (pre-)
- ▲ Quiz 3: “Preganglionic cell bodies at the sacral spinal level are a feature of the ____ division of the autonomic nervous system.” (parasympathetic)

The basic questions on the nervous system were intermingled with questions on regional anatomy. Students were told that the quizzes were in a fill-in-the-blank format, but they were not explicitly told there would be sometimes be repetition of information in the questions. Feedback was given at the end of each quiz by showing the correct answers for each of the quiz questions and whether the student’s response was correct or incorrect.

The quizzes were self-paced. Students took the second quiz on the same day and in the same session, roughly half an hour after the first. They took the third and final quiz on the questions a week later. This final post-test was given as a measure of long-term retention, and a one-week interval was chosen to be comparable to laboratory studies of repeated testing (e.g., Karpicke and Roediger, 2007; 2008).

Questions on the nervous system were graded by hand, and the names of the students were blacked out and replaced with a numerical identifier. The overall scores on the anatomy final examination and the scores on nervous system questions on the final examination were analyzed as well. In addition, the course evaluation included a question about the practice quizzes and provided opportunities for the students to make written comments. The practice quizzes were required for the Pre-entry Program, but the scores did not affect the final grade.

RESULTS

Two students did not attend the final testing session, so their data were eliminated from analyses. Both objective (quiz scores) and subjective (course evaluations) measures of learning and retention are reported here. With regard to subjective measures, unlike the optional practice quizzes taken by the first-year medical students in the pilot study, the mandatory practice quizzes were unpopular with students in the Pre-entry Program. Only 50% of the students thought that the quizzes were helpful or very helpful; the rest were undecided (12.5%) or else found them not helpful (37.5%). One problem with the Pre-entry Program quizzes is that Blackboard does not offer a spell-checker, and it was difficult to anticipate all of the possible written responses. As a result, responses that were spelled incorrectly or partially correct were scored as incorrect by Blackboard. It was clear that the students found it frustrating to receive lower scores than they deserved, even though the scores did not contribute to their grades. This problem was corrected by hand-grading the questions for the study so that the scores reported here accurately reflect all correct answers given by students.

All three written comments about the practice quizzes were negative. Comments included a student who said the quizzes “did not in any way demonstrate my knowledge of the relevant material. This was largely due to the fill-in-the-blank format.” and the two other students both commented that “the quizzes did not reflect the type of testing on the final.” It was clear that the students would have preferred quizzes with multiple choice questions in the USMLE format, such as those we use on the written examinations. In addition, one student listed the quizzes as a “major weakness” of the Human Anatomy course.

Nonetheless, although students were not enthusiastic about the quizzes, an analysis of the questions dealing with the nervous system showed that the students did, indeed, learn and remember the material and greatly improved their performance (Fig. 1). A repeated measures ANOVA revealed a highly-significant effect of testing ($F(2, 36) = 26.03$, $MSE = 163.24$, $P < 0.001$). As shown in Figure 1, student scores improved 8.7% on the second quiz, given just half an hour after the first, and scores were 29% higher on the third and final quiz compared to the first test. Paired t-tests revealed significant improvements in memory performance on each subsequent quiz, from Quiz 1 to Quiz 2 ($t(18) = 3.06$, $P < 0.01$) and Quiz 2 to Quiz 3 ($t(18) = 3.98$, $P < 0.001$). It is important to emphasize that the knowledge tested on the third quiz was retained for at least one week, the time that elapsed between taking the second and third quizzes.

There was also a good correlation ($r(19) = 0.51$, $P < 0.02$) between the scores on the third quiz and the overall scores on the final examination. The final examination score increased by 0.6 points, on average, when the practice quiz score increased by 1 point ($P = 0.005$). Another indication that the quizzes were helpful was that the students did better on the two questions on the final examination dealing with the nervous system than they did on the questions dealing with other topics. Although the overall average on the final examination was only 62%, 13 of the 21 students answered both questions about the nervous system correctly and all but three got at least one correct.

Discussion

Among students in the Pre-entry Program, there was a large, statistically significant increase in scores from the first to the third quizzes, 29%, on average. There was also a positive correlation between the quiz scores and the scores on the final examination. It is important to point out that the students were learning anatomy by many other means, not just the quizzes. After the initial lecture on the nervous system, there was more formal instruction about this topic, including lectures about the back and spinal cord as well as dissections of these regions. In addition, the teaching assistants conducted an informal review of the nervous system. A more rigorous test of the impact repeated testing, alone, as a tool for learning anatomy would require repeating the current study under more controlled laboratory conditions, similar to the research on vocabulary words by Karpicke and Roediger (2008), for example. On the other hand, this study was valuable as a demonstration of the “testing effect” in a classroom setting; scores increased by 8.7%, on average, on the second version of the quiz without any additional studying.

Only about half of the students in the Pre-entry Program found the quizzes helpful, but a much higher percentage, nearly 80%, of the first-year medical students found the quizzes helpful. There was an even more dramatic difference between the written comments from the two groups. There are many possible explanations for this discrepancy. The most likely explanation is that the first-year medical students appreciated tools that like the practice quizzes aid in learning the material, but the students in the Pre-entry Program did not realize that the quizzes were helpful. In other words, the medical students really are “memorizing the exact sentences in the syllabus”, as one of the students in the Pre-entry Program wrote in the course evaluation. Aside from the material in the clinical correlations, which was not included in the practice quizzes, the syllabus contains very little extraneous information, and the first-year students seemed to realize that. Another possibility is that the Pre-entry students, accustomed to doing very well on tests, might have found the experience of doing poorly on the quizzes aversive, even though the grades did not count. The first-year medical students might have been less concerned about this issue, again due to their more extensive experience with this and other medical school classes.

It is very likely that optional quizzes, available online at any time, would be more popular than mandatory quizzes on Friday afternoons. In order to obtain objective data about the effects of the practice quizzes, it was necessary to give them under controlled conditions. However, in the future, it may be feasible to simply ask the students to do their own work and not to look up the answers. Indeed, this approach might be preferable because it would provide other measures of quiz participation besides the scores. Web-based learning has been used successfully before in medical education (e.g., Marsh et al., 2008). Other changes that could be implemented easily and might make the experience of taking the practice quizzes more pleasant are to add a spell checker and to accept a wider range of correct answers. This problem did not affect the objective results because the nervous system questions were graded by hand, but it could have affected the subjective evaluations of the study. In future studies, students should have the option of correcting spelling errors during the quiz so that they receive a more accurate grade.

Smythe and Hughes (2008) showed that students in self-directed instruction who already feel overloaded with work are less likely to spend time on important, self-directed activities that could benefit learning. This is another possible reason why pre-entry students were more negative about the quizzes than medical students in the pilot study. Students in the Pre-entry Program have to learn a great deal of information in a short period of time over the summer. First-year medical students in the pilot study had the advantage of learning in a slightly more relaxed time frame of an entire semester. Pre-entry students may have felt that any additional time spent out of the classroom or laboratory was only adding to the burden

of information to be learned in a short time. Smythe and Hughes (2008) convincingly demonstrate that perceptions of difficulty, utility, and time investment are important in influencing how students feel about learning. In the future, it might be helpful to present the data from the current study to students in future studies, so they can see the empirical evidence that these practice quizzes have a substantial positive impact on both their final grades and their long-term retention of the information.

A change in the question format for the quizzes was suggested by some students in the Pre-entry Program. They requested multiple choice questions similar to those now used on the midterm and final examinations and also used on the United States Medical Licensing Examination. However, there are several reasons why this might not be advisable. There are many other sources for questions like these, including the syllabus and the textbooks required for the Human Anatomy course. While these questions are useful for assessment and for students who need to review anatomy, they are designed for students who have already mastered the material, not for beginning students. It would be possible for a bright first-year student to memorize the answers to frequently-asked questions and pass an examination without really understanding anatomy. In addition, memory research has demonstrated that deeper levels of learning occur when retrieval is more effortful and requires students to think through and generate their own answers, rather than choosing from a list of options (Gardiner et al., 1973; Bjork, 1994; Roediger and Karpicke, 2006b; Carpenter, 2009). Indeed, research suggests that the power of the testing effect “is not driven primarily by the similarity of initial and final tests but instead by retrieval processes taking place during the initial test” (Carpenter, 2009).

Although fill-in-the blank questions are relatively difficult to grade, they are very helpful for beginning students. Like questions on an anatomy practical examination, they require recall of the answer. Another advantage of fill-in-the blank questions is that it is possible for instructors to write a large number of easy questions, providing more opportunities for test-based learning. Relatively easy questions are particularly helpful when students first learn new material (Larsen et al., 2008). Marsh et al. (2008) found that more complicated learning modules were more beneficial towards the end of the course after more basic learning had taken place. Practice quizzes are just one of many methods that the students should use to help them learn anatomy.

Some educators may see incorporation of quizzes like these into the classroom as a drawback because they appear to focus on fact learning rather than comprehension of higher-order concepts and systems-level outlooks that are critical to clinical practice. Burns (2010) refers to the testing of small, easy-to-memorize amounts of information as anatomizing or “tidbitting” and eschews its use in medical classrooms because it is detrimental to higher-order learning. The repeated questions used in the current study did not directly test higher-order concepts that medical students are expected to learn in the course of their education. However, the current study used repeated testing solely as a learning aid. Moreover, an attempt was made to structure the quiz questions so they were not simple, flash-card facts. Questions frequently linked more than one related concept in order to foster these learning connections through exposure and practice. The focus of the paper by Burns (2010) is on assessment testing such as final examinations, which do, indeed, capture comprehension and understanding of higher order concepts. Burns (2010) makes an excellent argument that medical education will suffer if it supports learning-through-memorization, but it can be argued that all learning must have a solid foundation of facts that serves as a scaffold for higher order learning and linking of concepts. Ward and Walker (2008) reported that stronger students used multiple studying methods, whereas weaker students relied on a single method. Practice quizzes like the ones in the current study would help build and bolster such a foundation.

Another, unexpected benefit of the practice quizzes was that a student with a previously undiagnosed learning disability was identified. There was no time limit for the quizzes, but there was one student who consistently spent a great deal more time on them than the others had. When asked about this, he responded that the same had also been true in college. Testing by a clinical psychologist associated with the Pre-entry Program indicated that he would qualify for additional time on tests in medical school.

The practice quizzes will help meet two standards related to self-assessment opportunities for students issued by the Liaison Committee on Medical Education (LCME, 2010), the accrediting body for the medical school curriculum, including ED-5-A which states: “A medical education program must include instructional opportunities for active learning and independent study to foster the skills necessary for lifelong learning... These skills include self-assessment on learning needs... Medical students should receive explicit experiences in using these skills, and they should be assessed and receive feedback on their performance” (LCME, 2010) and ED-31 which states, “Each medical student in a medical education program should be assessed and provided with formal feedback early enough during each required course to allow sufficient time for remediation. Although a course...may not have sufficient time to provide a structured formative assessment, it should provide alternate means (e.g., self-testing, teacher consultation) that will allow medical students to measure their progress in learning” (LCME, 2010). Because the quizzes will be available on Blackboard, they will be available to the students at the time they determine is most useful to promote learning, and feedback will be automatically provided. As stated on some of the evaluations from the first study, students found the quizzes useful for identifying areas where they were potentially weak and for reinforcing concepts they already knew.

Practice quizzes such as these are expected to be useful in other courses in the future. The instructors from the University of Texas-Houston Dental School have already discussed the possibility of using them in Human Anatomy courses for dental and dental hygiene students. The results should also be applicable to other disciplines and to undergraduate students, such as those at Rice University. For example, preliminary results from this study were presented to professors there, including one from the Art Department. Even though there were clearly differences between the material taught in Art History and Human Anatomy and also between the two sets of students, many similarities became apparent. For example, descriptions of cathedrals featured prominently in the Art History curriculum, and the students were expected to learn the names of the various features. There were even terms in bold-face type in the Art History syllabus, as there are in the Human Anatomy dissector. The current study serves as an important demonstration that the repeated testing paradigm that has been used so successfully in the laboratory also offers many benefits in the medical school classroom. The findings also suggest that practice quizzes would be helpful in a variety of other classroom settings.

Acknowledgments

The authors wish to thank the Office of Educational Programs at the University of Texas Medical School-Houston for their support, Paul Coleman for assistance in coding and Dr. Alice Chuang for assistance with the statistical analysis. They are also very grateful to the instructors in the Human Anatomy course for giving the permission to write questions based on their material and to the students who answered those questions.

Grant sponsors: Teagle Foundation, Grant number: R03032; National Eye Institute, Core Grant number EY10608.

LITERATURE CITED

Bergman EM, Prince KJ, Drukker J, van der Vleuten CP, Scherpbier AJ. How much anatomy is enough? *Anat Sci Educ*. 2008; 1:184–188. [PubMed: 19177406]

Anat Sci Educ. Author manuscript; available in PMC 2012 September 1.

- Butler AC, Roediger HL III. Testing improves long-term retention in a simulated classroom setting. *Eur J Cogn Psychol.* 2007; 19:514–527.
- Butler AC, Roediger HL 3rd. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Mem Cognit.* 2008; 36:604–616.
- Bjork, RA. Memory and metamemory considerations in the training of human beings. In: Metcalfe, J.; Shimamura, AP., editors. *Metacognition: Knowing About Knowing.* 1st Ed.. MIT Press; Cambridge, MA: 1994. p. 185-205.
- Burns ER. “Anatomizing” reversed: Use of examination questions that foster use of higher order learning skills by students. *Anat Sci Educ.* 2010; 3:330–334. [PubMed: 21046570]
- Carrier M, Pashler H. The influence of retrieval on retention. *Mem Cognit.* 1992; 20:633–642.
- Carpenter SK. Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *J Exp Psychol Learn Mem Cogn.* 2009; 35:1563–1569. [PubMed: 19857026]
- Chan JC, McDermott KB, Roediger HL 3rd. Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *J Exp Psychol Gen.* 2006; 135:553–571. [PubMed: 17087573]
- Drake RL, McBride JM, Lachman N, Pawlina W. Medical education in the anatomical sciences: the winds of change continue to blow. *Anat Sci Educ.* 2009; 2:253–259. [PubMed: 19890982]
- Gardiner JM, Craik FI, Bleasdale FA. Retrieval difficulty and subsequent recall. *Mem Cognit.* 1973; 1:213–216.
- Glover JA. The “testing” phenomenon: Not gone but nearly forgotten. *J Educ Psychol.* 1989; 81:392–399.
- Karpicke JD, Roediger HL 3rd. The critical importance of retrieval for learning. *Science.* 2008; 319:966–968. [PubMed: 18276894]
- Karpicke JD, Roediger HL 3rd. Is expanding retrieval a superior method for learning text materials? *Mem Cognit.* 2010; 36:116–124.
- Larsen DP, Butler AC, Roediger HL 3rd. Test-enhanced learning in medical education. *Med Educ.* 2008; 42:959–966. [PubMed: 18823514]
- LCME. Functions and Structure of a Medical School: Standards for Accreditation of Medical Education Programs Leading to the M.D. Degree. Liaison Committee on Medical Education; Washington, DC: June. 2010 p. 282010Liaison Committee on Medical EducationURL: <http://www.lcme.org/functions2010jun.pdf>
- Logan JM, Balota DA. Expanded vs. equal interval spaced retrieval practice: exploring different schedules of spacing and retention interval in younger and older adults. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn.* 2008; 15:257–280. [PubMed: 18421627]
- Marsh KR, Giffin BF, Lowrie DJ Jr. Medical student retention of embryonic development: impact of the dimensions added by multimedia tutorials. *Anat Sci Educ.* 2008; 1:252–257. [PubMed: 19109854]
- McDaniel MA, Anderson JL, Derbish MH, Morrisette N. Testing the testing effect in the classroom. *Eur J Cogn Psychol.* 2007; 19:494–513.
- Roediger HL III, Karpicke JD. Test-enhanced Learning: Taking memory tests improves long-term retention. *Psychol Sci.* 2006a; 17:249–255. [PubMed: 16507066]
- Roediger HL III, Karpicke JD. The power of testing memory: Basic research and implication for educational practice. *Perspect Psychol Sci.* 2006b; 1:181–210.
- Roediger, HL., III; Agarwal, PK.; Kang, SHK.; Marsh, EJ. Benefits of testing memory: Best practices and boundary conditions. In: Davies, GM.; Wright, DB., editors. *Current Issues in Applied Memory Research.* 1st Ed.. Psychology Press; Brighton, UK: 2010. p. 13-49.
- Smythe G, Hughes D. Self-directed learning in gross human anatomy: Assessment outcomes and student perceptions. *Anat Sci Educ.* 2008; 1:145–153. [PubMed: 19177401]
- Sugand K, Abrahams P, Khurana A. The anatomy of anatomy: A review for its modernization. *Anat Sci Educ.* 2010; 3:83–93. [PubMed: 20205265]
- Thompson CP, Wenger SK, Bartling CA. How recall facilitates subsequent recall: A reappraisal. *J Exp Psychol Hum Learn Mem.* 1978; 4:210–221.

Ward PJ, Walker JJ. The influence of study methods and knowledge processing on academic success and long-term recall of anatomy learning by first-year veterinary students. *Anat Sci Educ*. 2008; 1:68–74. [PubMed: 19177384]

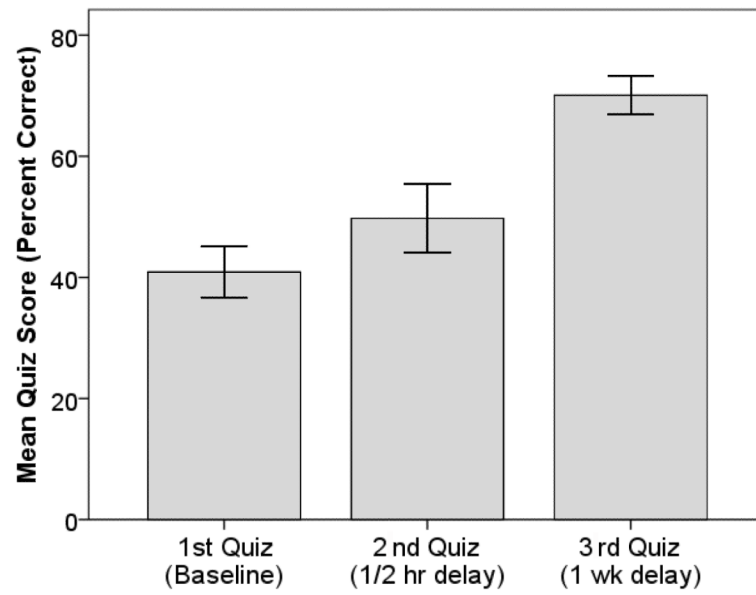


Figure 1. Mean quiz scores given as percent correct on each of three quizzes given after one-half hour and one week. Error bars = ± 1 standard error of the mean.