2-17-2015

The effects of test-enhanced learning on long-term retention in AAN annual meeting courses.

Douglas P Larsen
Department of Neurology, Washington University in St. Louis School of Medicine

Andrew C Butler
Department of Psychology and Neuroscience, Duke University

Wint Y Aung
Department of Neurology, Washington University in St. Louis School of Medicine

John R Corboy
Department of Neurology, University of Colorado School of Medicine

Deborah I Friedman
Department of Neurology, University of Texas Southwestern Medical School

See next page for additional authors

Let us know how access to this document benefits you

Follow this and additional works at: https://jdc.jefferson.edu/neurologyfp

Part of the Neurology Commons

Authors

Douglas P Larsen, Andrew C Butler, Wint Y Aung, John R Corboy, Deborah I Friedman, and Michael R
Sperling

The effects of test-enhanced learning on long-term retention in AAN annual meeting courses

Douglas P. Larsen, MD,
    MEd
Andrew C. Butler, PhD
Wint Y. Aung, BA
John R. Corboy, MD
Deborah I. Friedman,
    MD, MPH
Michael R. Sperling, MD

Correspondence to
Dr. Larsen:
larsend@neuro.wustl.edu

ABSTRACT

Objective: We measured the long-term retention of knowledge gained through selected American Academy of Neurology annual meeting courses and compared the effects of repeated quizzing (known as test-enhanced learning) and repeated studying on that retention.

Methods: Participants were recruited from 4 annual meeting courses. All participants took a pretest. This randomized, controlled trial utilized a within-subjects design in which each participant experienced 3 different postcourse activities with each activity performed on different material. Each key information point from the course was randomized in a counterbalanced fashion among participants to one of the 3 activities: repeated short-answer quizzing, repeated studying, and no further exposure to the materials. A final test covering all information points from the course was taken 5.5 months after the course.

Results: Thirty-five participants across the 4 courses completed the study. Average score on the pretest was 36%. Performance on the final test showed that repeated quizzing led to significantly greater long-term retention relative to both repeated studying (55% vs 46%; $t[34] = 3.28$, SEM = 0.03, $p = 0.01$, $d = 0.49$) and no further exposure (55% vs 44%; $t[34] = 3.16$, SEM = 0.03, $p = 0.01$, $d = 0.58$). Relative to the pretest baseline, repeated quizzing helped participants to retain almost twice as much of the knowledge acquired from the course compared to repeated studying or no further exposure.

Conclusions: Whereas annual meeting continuing medical education (CME) courses lead to long-term gains in knowledge, when repeated quizzing is added, retention is significantly increased. CME planners may consider adding repeated quizzing to increase the impact of their courses. *Neurology*® 2015;84:748–754

GLOSSARY

AAN = American Academy of Neurology; **ANOVA** = analysis of variance; **CME** = continuing medical education.

Continuing medical education (CME) comprises a major component of both maintenance of certification and maintenance of licensure in most states.[1,2] The rationale behind this position is that life-long learning is a critical part of professional development.[2] The American Academy of Neurology (AAN) is a major source of CME for neurologists in the United States. Each year, thousands of neurologists spend considerable time and money to attend the annual meeting of the AAN in order to meet their CME requirements and to learn about the latest advances in the field. In order to better meet the educational needs of practicing neurologists, the AAN has developed interactive education sessions such as NeuroFlash and Morning Report sessions.[3] NeuroFlash sessions use audience response systems to engage learners in presentations about recent developments in a topic. Morning Reports use a case-based discussion format to explore challenging clinical scenarios. These new formats focus not only on facts but on the clinical application of those facts, which is a key part of the definition of competence used for accreditation of CME courses.[4,5] Although increasing interaction during these sessions is a good educational strategy, a single, brief exposure to information may not be enough to produce

robust long-term retention over time.[6] This newly acquired knowledge must be maintained in order for these courses to achieve the goal of producing enduring changes in competence.

One possible intervention that may improve retention of the knowledge gained from these courses is test-enhanced learning. Test-enhanced learning is a concept based on the finding that retrieving information from memory increases retention of that information over time.[7,8] Although memory retrieval is often thought of as a neutral event, the act of retrieving information from memory actually changes memory by strengthening it.[9] Activities that incorporate retrieval practice, such as tests or quizzes, are potent learning tools that can produce robust long-term retention. Studies using medical students, residents, middle school students, and older adults have all shown benefits in knowledge retention using this technique.[6,10–14] However, the effects of test-enhanced learning have not been investigated in CME settings.

Another possible intervention for promoting retention after CME courses is to simply provide additional exposure to information through repeated study. Several studies of CME have shown that multiple exposures to information increases retention relative to a single exposure.[15] Repeated study may also be easier to implement because generating activities to provide retrieval practice (e.g., quizzes and tests) after CME involves substantial investments of time. In order to assess the benefits and costs of these 2 potential interventions, we sought to directly compare them. In this study, we investigated how repeated retrieval via quizzes vs repeated studying of the material affects the long-term retention of knowledge gained through CME courses at the AAN annual meeting. We also included a control condition in which no further exposure to the material was provided in order to assess how much of it would be retained without any postcourse intervention.

METHODS Courses and participants. Participants were recruited from 4 courses presented during the 2012 AAN annual meeting held in New Orleans, Louisiana. The 4 courses were NeuroFlash Epilepsy, NeuroFlash Child Neurology, Morning Report Multiple Sclerosis, and Morning Report Challenging Headache Cases. The courses were chosen because they had high levels of attendance in previous years and represented a variety of topics. We anticipated that based on these attendance trends we would be able to recruit an adequate number of participants to achieve meaningful results based on the sample sizes required in previous studies.[6,10,11] All preregistered course participants were eligible for the study. If participants had previously opted out of receiving surveys from SurveyMonkey.com or did not have a valid e-mail address, they were not included in the number of individuals invited to participate in the study.

Standard protocol approvals, registrations, and participant consents. Preregistered participants were invited by e-mail to participate prior to the course. If participants chose to enroll in the study, an information screen was provided explaining the study and informing participants that the choice to participate constituted consent. Participants were offered the opportunity to be included in a drawing for a free course from the AAN as compensation for their participation. Six free courses were offered by drawing for each course in the study. Due to law, only legal residents of the United States could be included in the drawings. Nonresidents of the United States had to waive their participation in the drawing in order to participate. The study was approved by the Washington University in St. Louis Institutional Review Board.

Materials. For each course, the instructor and the first author identified the key information points from the material that would be taught to participants. The number of key information points covered in a course ranged between 36 and 58. Using these information points, a set of short-answer questions and corresponding answers was created for each of the 4 courses. The questions were used to construct a pretest, learning quizzes, and a final test for each course. The answers to the questions were used to provide feedback on the learning quizzes. For each question and answer, a corresponding review statement was created that presented the information points as facts to be studied. The review statements covered the same material as the questions and answers; the only difference was that the question and answer were rewritten to read as a statement. An example short-answer question from the multiple sclerosis course was "What are the 3 characteristics of MRI findings on an initial scan that satisfy the criteria for dissemination in time?" Participants then were required to list those characteristics. The corresponding review item was simply a statement of the 3 characteristics of MRI findings on an initial scan that satisfied the criteria for dissemination in time. All materials were delivered to participants via SurveyMonkey.com. Once the pretests were scored, the set of questions was divided into thirds to create 3 practice quizzes. The questions were stratified by how frequently they were answered correctly on the pretest so that each of the follow-up learning quizzes was roughly the same level of difficulty. The final test was identical to the pretest.

Procedures. All individuals who were registered for each of the 4 courses were invited to participate in the study via e-mail. Individuals who chose to participate accessed the pretest via a Web link in the invitation e-mail. All pretests were scored prior to the course. Materials for each course were divided into thirds in a stratified fashion based on the pretest score so that each third was relatively equivalent in difficulty. The items were then randomized in a counterbalanced fashion to be learned by each participant in 1 of 3 follow-up activities: repeated testing as a quiz item, repeated study as a review item, or no further exposure. The counterbalancing ensured that each third of the material was equally assigned to all 3 learning activities across

participants. Participants were also stratified based on their pretest scores to ensure that mean pretest performance was close to equivalent across the different follow-up activities for each third of the material. Using this within-subjects, counterbalanced design, each participant took part in all 3 of the follow-up activities and each item was learned in each of the 3 different activities.

Immediately after the course was completed, participants received a Web link by e-mail to the initial set of follow-up activities. Participants first completed the learning quiz, then they studied the review statements for the material assigned to repeated study, and finally they received the answers to the questions on the learning quiz as feedback. Participants had approximately a week to complete the follow-up activities. They received 3 more quizzes and review statements at weekly intervals. The material covered in the quizzes and review statements was the same each week, but the order of presentation of the information points was different from week to week. Approximately 5.5 months after the course, all participants took the final test, which covered all of the course material.

Scoring and analysis. Each item on the pretests and final tests was scored as correct or incorrect by 2 individuals (D.P.L. and W.Y.A.). Differences were resolved by discussion until a consensus was reached. The kappa statistic measuring interrater reliability was high ($\kappa = 0.94$). Given the high rate of agreement, the practice quizzes were scored by a single individual (W.Y.A.). The results from all 4 courses were combined for statistical analysis. Final test scores were analyzed using a repeated-measures analysis of variance (ANOVA) followed by planned post hoc comparisons using paired-samples t tests. A Bonferroni correction was used to correct for multiple comparisons when analyzing the final test results. Effect size was determined using η^2 for the ANOVA and Cohen d for the t tests. Statistical analyses were performed using SPSS 21 software (SPSS, Chicago, IL).

RESULTS Participants. A total of 251 individuals registered for the 4 CME courses and were invited to take the pretest beforehand (see table 1 for values divided by course). An additional 20 individuals who were preregistered had either opted out of receiving surveys from SurveyMonkey.com or did not have a valid e-mail address and were therefore not counted in the number of individuals who received the invitation because they were not contactable. Ninety-six individuals completed the pretest and were randomized to the various conditions. Forty-one individuals initiated the final test. Three individuals

initiated the final test but did not complete it and 3 individuals completed the final test but had not taken any of the follow-up learning quizzes; these 6 individuals were not included in the data analysis. The remaining 35 individuals who had both completed the final test and had completed at least one set of follow-up activities (i.e., learning quiz and review statements) were included in data analysis. Of these participants, 20 completed all 4 sets of follow-up activities, 6 completed 3, 6 completed 2, and 3 completed 1.

Pretest scores, initial learning, and review frequency. The mean pretest score across all 4 courses was 36% (see table 2 for scores by course), indicating that participants had relatively little knowledge of the content covered beforehand. Mean score on the initial learning quiz across all 4 courses was 57%, which suggests that participants gained a substantial amount of knowledge from the CME courses. Although each follow-up learning quiz covered only one-third of the material for the course, the average across all of the initial quizzes provides a rough estimate of the level of retention soon after completion of the course.

To assess the degree of engagement of participants in studying the review items as well as studying the answers to the quizzes, they were asked at the end of each set of follow-up activities how many times they had read through the review items and the answers to the quiz items. On average, participants read through the review items 1.90 times and read through the answers to the quiz items 1.54 times (see table 3 for frequency of review by course). Of note, participants in the Challenging Headache Cases course and the Child Neurology course tended to read through the review items and the quiz answers fewer times than the other 2 courses.

Final test scores. Final test scores were analyzed as a function of the learning activity to which each portion of the material was assigned. Participants scored an average of 55% correct on questions about material that was learned through repeated quizzing (see table 2 for results by course). For material that was learned through repeatedly studying the review statements, participants scored an average of 46% correct. For material that participants had no further exposure to after the course, they scored an average of 44% correct (figure). A repeated-measures ANOVA showed a main effect of learning condition ($F_{2,68} = 7.31$, SEM $= 0.02$, $p = 0.001$, $\eta^2 = 0.18$). The planned post hoc comparisons indicated that the level of performance produced by repeated quizzing was significantly greater than both repeated studying ($t[34] = 3.28$, SEM $= 0.03$, $p = 0.01$, $d = 0.49$) and no further exposure ($t[34] = 3.16$, SEM $= 0.03$, $p = 0.01$, $d = 0.58$). All 3 learning activities produced

Table 1   Number of participants by course

Course	Invited	Pretest	Final
Epilepsy	112	43	15
Multiple sclerosis	79	28	13
Headache	42	17	4
Child neurology	18	8	3
Total	251	96	35

Invited = number of individual who were registered for the course and invited to participate in the study. Pretest = number of individuals who completed the pretest. Final = number of individuals who completed the final test and at least one set of follow-up activities (learning quiz/review statements).

Course	Repeated quizzing			Repeated study			No further exposure		
	Pretest	Final	Difference	Pretest	Final	Difference	Pretest	Final	Difference
Epilepsy (n = 15)	36 (13)	57 (20)	21 (21)	35 (12)	43 (16)	8 (16)	33 (13)	45 (20)	12 (23)
Multiple sclerosis (n = 13)	38 (15)	54 (21)	16 (22)	47 (16)	51 (16)	4 (18)	41 (13)	48 (16)	7 (10)
Headache (n = 4)	44 (16)	61 (7)	17 (18)	32 (15)	46 (6)	14 (12)	24 (13)	37 (8)	13 (19)
Child neurology (n = 3)	26 (19)	41 (23)	15 (19)	32 (8)	44 (14)	12 (21)	30 (3)	36 (5)	6 (6)
Average (n = 35)	37 (15)	55 (19)	19 (18)	39 (15)	46 (15)	7 (17)	35 (13)	44 (17)	9 (17)

scores that were significantly greater than the corresponding pretest scores; however, repeated quizzing produced a larger gain in knowledge ($t[34]$ = 5.44, SEM = 0.03, p = 0.000005, d = 0.95) relative to repeated study ($t[34]$ = 2.83, SEM = 0.03, p = 0.008, d = 0.51) and no further exposure ($t[34]$ = 3.17, SEM = 0.03, p = 0.003, d = 0.60).

DISCUSSION The AAN has invested significant time and money in developing CME courses. Our study shows that these courses address important needs of its members. The courses were effective in producing short-term knowledge gains. Participants only were able to answer 36% of questions correctly on the pretest. The scores on the initial learning quizzes, which approximate the immediate effect of the course on learning, showed a substantial increase in knowledge, with participants answering 57% of the questions correctly. Although higher scores might have been expected so soon after taking the course, this finding is generally consistent with the rate of short-term retention in lecture-based courses.[6,16] The direct comparison of the initial learning quiz performance with the pretest should be interpreted with caution because each participant only reanswered some of the questions; that said, all participants took the same course and when combined across participants the initial quiz questions cover all of the course material.

With respect to long-term retention, our study showed that some of the knowledge gained from the CME courses was retained 5.5 months later even without any further exposure. Final test performance on the material for which participants had no further exposure after the course was 44%, which represents an increase when compared to the pretest performance. However, the degree of this learning is probably less than most

CME organizers would hope to achieve. If the initial learning quiz performance is used as a benchmark for the postcourse level of knowledge, then some of the material was forgotten over time. This finding is consistent with other studies that have shown similar forgetting effects.[6]

We found that additional exposure to the course materials through repeated studying led to a similar level of retention as having no further exposure after the course (final test scores of 46% compared with 44%). Although this finding is different than some CME studies that have shown a benefit of repeated exposure, it is consistent with findings of other studies that have compared repeated studying and repeated testing.[9] Continuing to repeatedly study once the information is initially learned may not be an effective learning strategy.
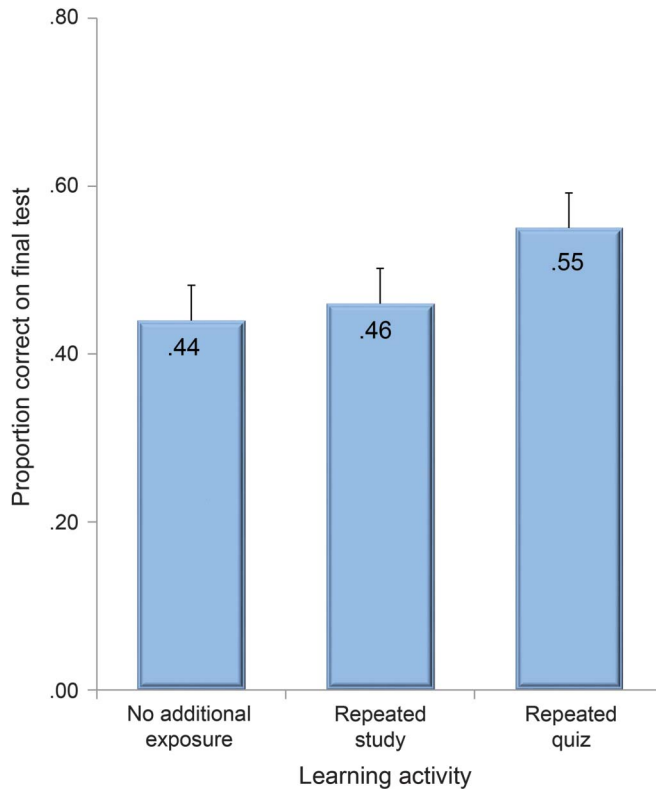
On the other hand, repeated quizzing of the course material produced much better long-term retention relative to repeated studying or no further exposure. The improvement from the pretest to the final test was almost twice as large for repeated quizzing compared to the 2 other learning activities. In addition, we found very little forgetting of material that had been repeatedly quizzed from the initial learning quiz to the final test. These findings are consistent with a growing body of evidence that test-enhanced learning consistently produces superior retention over the long run.[6,10–14]

The test-enhanced learning literature suggests that to maximize the mnemonic effects of testing, the following 5 principles should be followed: (1) tests should arise from educational objectives, (2) questions should require the generation of answers rather than the recognition of answers, (3) testing should allow for repeated retrieval practice, (4) repeated tests should be adequately spaced so as to require effortful

	Epilepsy	Multiple sclerosis	Headache	Child neurology	All participants
Review items	2.00	2.09	1.58	1.17	1.90
Quiz answers	1.53	1.77	1.26	1.25	1.54

Figure   Total percent correct on the final test collapsed across courses for each learning activity

it also has some limitations. One limitation is that there was considerable attrition of participants over time, which led to a relatively small sample size. Ninety-seven individuals took the pretest, but only 35 individuals completed at least one of the practice quizzes and the final test. Many of those who completed the final test did not complete all 4 practice quizzes. The large attrition rate may limit the generalizability of our results because only the most motivated individuals may have completed the study. Nevertheless, the fact that a significant effect was obtained with such a small sample and without everyone completing all practice quizzes speaks to the potentially powerful effects of test-enhanced learning. Our results may underestimate the potential effects that repeated testing could have if individuals were more motivated to engage in postcourse learning activities. However, our study must be considered exploratory given the small numbers.

Another limitation of our study was that it included only 4 CME courses. Given this relatively small selection of courses, it is possible that our results might not apply to all courses offered. However, it should be pointed out that the courses used in the study varied greatly in terms of subject matter and the instructor's method of teaching. In addition, the extensive literature on test-enhanced learning has shown that the benefits of practicing retrieval hold across many different types of material and settings. Thus, our results would indicate that other CME courses could benefit from the incorporation of test-enhanced learning principles.

The retention of information measured on our quizzes and final tests is similar to other studies that have measured the retention of lecture-type courses.[6,16] However, it should be noted that question format and difficulty of the questions may influence how high or low the absolute percentage correct may be for any given quiz in a course. More difficult or detailed questions may result in lower scores. Despite this fact, our findings indicate that repeated quizzing appears to be beneficial even when questions are difficult. Also, while the pattern of our results was consistent across courses, the degree of benefit from repeated quizzing was variable. Both the Challenging Headache Cases course and the Child Neurology course showed a small benefit to repeated quizzing. Both of these courses had very small numbers of participants who completed the study, making firm conclusions difficult to draw. The relative difficulty of the materials in various courses may have played a role as well as the degree of engagement with those materials. Participants in these courses also tended to read the answers to quizzes and the review items fewer times than participants in other courses, indicating that they may not have put as much effort into the learning process and many not have received as much of a benefit.

retrieval, and (5) feedback should be provided after each test.[17] Our testing regimen incorporated these principles. Questions were generated directly from the course materials, and instructors reviewed the questions to ensure that the important ideas were captured. Our tests used short-answer questions that required participants to generate the answers from memory. Although recognition questions have improved retention in some studies, other studies have found that test formats that require the more effortful generation of answers produce even better retention.[18,19] Rather than take a single test, participants in our study took 4 learning quizzes on a weekly basis. Prior studies have shown that testing has a dose effect and that repeated tests are better than single tests.[20] In addition, spacing repeated tests over weeks and months is likely to produce better retention over months and years than if tests are grouped more closely together.[21] Finally, the act of testing alone improves retention even without feedback, but feedback greatly accentuates the mnemonic benefits of testing.[22,23] In our study, participants received the answers to the questions as feedback after they had completed the quiz and studied the review statements about the other material. This approach is consistent with findings that delaying feedback further enhances its effects on learning.[24]

Our study demonstrates the potential benefits of using test-enhanced learning in a CME setting, but

A final limitation of our study was that we used the same questions for the pretest, the practice quizzes, and the final test. Thus, we focused on the retention of knowledge rather than the ability to apply this knowledge in different contexts (i.e., transfer of learning). Demonstrating that participants in a CME course can retain at least the basic facts presented in the course is an important first step in investigating test-enhanced learning in CME. The information presented and retained in these courses was clinically oriented and applicable. Other studies have shown that knowledge retained through repeated testing can be applied to novel circumstances.[25,26] Therefore, we anticipate that the additional knowledge retained by course participants through repeated quizzing should benefit their clinical practices.

Physicians spend considerable time and money taking CME courses such as those offered at the AAN annual meeting. In order to optimize physicians' return on their investment in these courses, organizers should use evidence-based interventions to increase long-term retention. This exploratory study may offer some suggestions for how learning in CME courses can be optimized. Our study shows that participants acquired knowledge from taking the CME courses that we studied, but much of this knowledge was forgotten over time, resulting in a small overall gain. Repeatedly studying the material from the course did not confer much additional benefit. However, repeated quizzing of the material helped physicians to retain more of the knowledge that they learned from the course. CME organizers may consider implementing repeated, spaced quizzes after their courses in order to optimize retention—the resulting increase in long-term retention of knowledge could make a difference in clinical practice.

REFERENCES

1. American Board of Psychiatry and Neurology. Maintenance of certification in neurology/child neurology. Available at: http://www.abpn.com/moc_neuro.asp. Accessed March 20, 2014.
2. Chaudhry HJ, Cain FE, Staz ML, Talmage LA, Rhyne JA, Thomas JV. The evidence and rationale for maintenance of licensure. J Med Regul 2013;99:19–26.
3. American Academy of Neurology. New Morning Report, NeuroFlash programs offer interactive learning opportunities. Available at: http://patients.aan.com/news/?event=read&article_id=8694. Accessed March 19, 2014.
4. Accreditation Council on Continuing Medical Education. What is the ACCME's definition of "competence" as it relates to the Accreditation Criteria? Available at: http://www.accme.org/ask-accme/what-accmes-definition-competence-it-relates-accreditation-criteria. Accessed March 20, 2014.
5. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990;65(9 suppl):S63–S67.
6. Larsen DP, Butler AC, Roediger HL III. Repeated testing improves long-term retention relative to repeated study: a randomized, controlled trial. Med Educ 2009;43:1174–1181.
7. Roediger HL III, Butler AC. The critical role of retrieval practice in long-term retention. Trends Cogn Sci 2011;15:20–27.
8. Larsen DP, Butler AC, Roediger HL III. Test-enhanced learning in medical education. Med Educ 2008;42:955–966.
9. Karpicke JD, Roediger HL III. The critical importance of retrieval for learning. Science 2008;319:966–968.
10. Larsen DP, Butler AC, Lawson AL, Roediger HL III. The importance of seeing the patient: test-enhanced learning with standardized patients and written tests improves clinical application of knowledge. Adv Health Sci Educ Theory Pract 2013;18:409–425.
11. Larsen DP, Butler AC, Roediger HL III. Comparative effects of test-enhanced learning and self-explanation on long-term retention. Med Educ 2013;47:674–682.
12. Kromann CB, Jensen ML, Ringsted C. The testing effect on skills might last 6 months. Adv Health Sci Educ Theory Pract 2010;15:395–401.
13. McDaniel MA, Agarwal PK, Huelser BJ, McDermott KB, Roediger HL III. Test-enhanced learning in a middle school science classroom: the effects of quiz frequency and placement. J Educ Psychol 2011;103:399–414.
14. Troyer AK, Hafliger A, Cadieux MJ, Craik FIM. Name and face learning in older adults: effects of level of processing, self-generation, and intention to learn. J Gerontol B Psychol Sci Soc Sci 2006;61:P67–P74.
15. Marinopoulos SS, Dorman T, Ratanawongsa N, et al. Effectiveness of Continuing Medical Education: Evidence Report/Technology Assessment No. 149 (prepared by the Johns Hopkins Evidence-Based Practice Center, under contract no. 290-02-0018): AHRQ Publication No. 07-E006. Rockville, MD: Agency for Healthcare Research and Quality; 2007.
16. Jones HE. The effects of examination on the performance of learning. Arch Psychol 1923/1924;10:1–70.
17. Larsen DP, Butler AC. Test-enhanced learning. In: Walsh K, ed. Oxford Textbook of Medical Education. Oxford: Oxford University Press; 2013:443–452.

18. Marsh EJ, Roediger HL III, Bjork RA, Bjork EL. The memorial consequences of multiple-choice testing. Psychon Bull Rev 2007;14:194–199.
19. Glover JA. The "testing" phenomenon: not gone but nearly forgotten. J Educ Psychol 1989;81:392–399.
20. Rawson KA, Dunlosky J. Optimizing schedules of retrieval practice for durable and efficient learning: how much is enough? J Exp Psychol Gen 2011;140:283–302.
21. 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: 354–380.
22. Butler AC, Karpicke JD, Roediger HL III. Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. J Exp Psychol Learn Mem Cogn 2008;34:918–928.
23. Pyc MA, Rawson KA. Why is test-restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. J Exp Psychol Learn Mem Cogn 2012;38:737–746.
24. Butler AC, Roediger HL III. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. Mem Cogn 2008;36:604–616.
25. Butler AC. Repeated testing produces superior transfer of learning relative to repeated studying. J Exp Psychol Learn Mem Cogn 2010;36:1118–1133.
26. Baghdady M, Carnahan H, Lam EWN, Woods NN. Test-enhanced learning and its effect on comprehension and diagnostic accuracy. Med Educ 2014;48:181–188.