

Test-enhanced learning in medical education

Douglas P Larsen,¹ Andrew C Butler² & Henry L Roediger III²

CONTEXT In education, tests are primarily used for assessment, thus permitting teachers to assess the efficacy of their curriculum and to assign grades. However, research in cognitive psychology has shown that tests can also directly affect learning by promoting better retention of information, a phenomenon known as the testing effect.

COGNITIVE PSYCHOLOGY RESEARCH Cognitive psychology laboratory studies show that repeated testing of information produces superior retention relative to repeated study, especially when testing is spaced out over time. Tests that require effortful retrieval of information, such as short-answer tests, promote better retention than tests that require recognition, such as multiple-choice tests. The mnemonic benefits of testing are further enhanced by feedback, which helps students to correct errors and confirm correct answers.

APPLICATION TO MEDICAL EDUCATION Medical educational research has focused extensively on assessment issues. Such assessment research permits the conclusion that clinical expertise is founded on a broad fund of knowledge and effective memory networks that allow easy access to that knowledge. Test-enhanced learning can potentially strengthen clinical knowledge that will lead to improved expertise.

CONCLUSIONS Tests should be given often and spaced out in time to promote better retention of information. Questions that require effortful recall produce the greatest gains in memory. Feedback is crucial to learning from tests. Test-enhanced learning may be an effective tool for medical educators to use in promoting retention of clinical knowledge.

KEYWORDS *education, medical; educational measurement/*methods; psychology/*education; Missouri; feedback; cognition; teaching/methods; randomised controlled trial [publication type].

Medical Education 2008; **42**: 959–966
doi:10.1111/j.1365-2923.2008.03124.x

INTRODUCTION

Medical students have proven themselves time and again as outstanding test takers by the time they reach medical school. Tests in education are almost exclusively used as assessment tools. Tests in the classroom measure what students have learned in the course and permit the instructor to rank order them for assigning grades, whereas standardised tests measure a student's aptitude for learning (e.g. college entrance examinations), knowledge of a specific area (e.g. specialty board certification) or general intelligence (e.g. the Wechsler Adult Intelligence Scale).¹

Although tests certainly have a function as assessment devices, we make the argument here that tests can also promote learning by directly increasing the retention of information. The fact that tests can be used to improve learning is not widely realised within education, especially higher education. In prior work, we have advocated an approach we call test-enhanced learning.^{2,3} The use of tests can have both direct and indirect effects in education. Both are important, although our research on test-enhanced learning focuses more on the direct effects of testing. We discuss both types of effect briefly.

The direct effect of testing is based on research showing that when students are tested on material, they remember that material much better than when they are not tested on the material. This is called the *testing effect*, and it holds true across a wide variety of materials and experimental conditions (see Roediger and Karpicke⁴ for a review). In fact, after reading

¹Department of Neurology, Washington University in St Louis, St Louis, Missouri, USA

²Department of Psychology, Washington University in St Louis, St Louis, Missouri, USA

Correspondence: Douglas P Larsen MD, Department of Neurology, Washington University School of Medicine, 660 South Euclid Avenue, Campus Box 8111, St Louis, Missouri 63110, USA.
Tel: 00 1 314 454 6120; Fax: 00 1 314 454 2523;
E-mail: larsend@neuro.wustl.edu

Overview

What is already known on this subject

In medical education, tests are used almost exclusively for assessment.

What this study adds

Research in cognitive psychology has shown that the use of tests may promote better retention of information. This is known as test-enhanced learning. To improve memory, tests should be used often but should be administered over time. Tests that require effortful recall (e.g. short-answer or essay tests) produce greater gains than do multiple-choice tests. Feedback is an important part of test-enhanced learning.

Suggestions for further research

Further research will be needed to validate the use of tests to promote the retention of information in specific educational settings.

material or hearing a lecture about a topic, being tested on the material provides a greater boost to later retention than does rereading the material. Being tested is a better way to learn material than is further study, as we discuss below.

The indirect effects of testing refer to the increase in the amount of study time and improvement in study strategies that result from frequent testing. Students tend to study material most thoroughly shortly before a test. If students are given only a midterm test and a final examination (as often happens in large university courses), they will probably have only two periods of intense study during the semester. By contrast, if students are tested frequently during a course (say, weekly), they will potentially be more likely to keep up with readings and space out their periods of study. Much research shows that spaced study sessions aid memory performance.^{5,6} In addition, if students test themselves as a strategy for learning, they can discover their own areas of weakness and re-study material in a purposeful way. These indirect benefits of frequent testing have been shown to improve performance in large lecture classes.^{7,8} Of course, the indirect benefits occur in tandem with the direct benefits of better retention of the tested material.

In this paper, we will first provide a brief review of the positive effects of testing on later retention as studied by cognitive psychologists, usually (but not always) in controlled laboratory experiments. Secondly, we then consider the theoretical framework of testing in medical education and how this has evolved over time. In the final section of the paper, we discuss how testing could be used as a memory tool in medical education and how test-enhanced learning fits into the existing literature of medical education assessment.

TEST-ENHANCED LEARNING IN COGNITIVE PSYCHOLOGY

Throughout the history of psychology, most theories of human learning have implicitly incorporated the assumption that learning occurs when people are exposed to material during study and, as a corollary, that testing represents a neutral event that permits an assessment of what was learned during study. Indeed, the idea that taking a test assesses the contents of memory without changing them is widespread, especially in education. This assumption was challenged by the results of several early studies that showed a mnemonic benefit of testing.^{9–11} Later research also showed that learning does occur during tests and led to the re-conceptualisation of retrieval as a memory modifier.^{12,13} Yet, despite substantial evidence that testing enhances retention, the potential for using tests to promote learning was largely overlooked during the 20th century, leading one educational psychologist to entitle his paper ‘The “testing” phenomenon: not gone, but nearly forgotten’.¹⁴ Fortunately, interest in the testing effect has increased markedly in recent years, producing a flurry of research articles that explore the ways testing can be used to increase retention.

Many different theories have been put forth over the years to explain the testing effect.⁴ For simplicity’s sake, these theories can be separated into two categories that roughly correspond with either side of the debate described above about whether learning occurs during study or testing. One category of theories revolves around the idea that the act of retrieving information from memory strengthens memory for that information and thus leads to better long-term retention. By contrast, the other category of theories suggests that testing is beneficial merely because it involves additional exposure to the material (i.e. another chance to study), an idea called the ‘total time hypothesis’. The total time hypothesis emerged from a context in which many testing effect

experiments compared a group that studied information and then took a test with a control group that studied the material once and performed no further activity (see Roediger & Karpicke⁴ for discussion). Because total processing time was not equated in these early experiments, the superior retention of the tested group could be attributed to the benefit of an additional exposure to the material rather than to the taking of a test. Ultimately, the total time hypothesis was found to be incorrect by experiments showing that taking a test leads to better retention than re-studying the material for an equivalent amount of time.^{2,14,15} As a result, theoretical explanations for the testing effect now focus on retrieval as the critical mechanism that underlies improved retention. One of the explanations for this phenomenon is that retrieval during a test involves active processing. Because a final test or application task will also involve active retrieval of information, the active retrieval in the initial testing essentially practises the skill that will be needed later. Studying the information, by contrast, does not involve practising the critical retrieval skills that will be required on a final test; hence the superiority of repeated testing to repeated rereading of information for long-term retention.⁴

To further illustrate the relative contributions of study and testing to learning, consider a simple, yet powerful, experiment reported by Karpicke and Roediger.¹⁶ The experiment investigated the optimal method for learning foreign language vocabulary with flash cards. In the first phase, students studied a list of 40 Swahili–English word pairs (e.g. *mashua–boat*) and were then tested on the whole list (e.g. *mashua–?*). Students engaged in this initial phase until they had recalled all 40 words correctly at least once. For the second phase, they were randomised to four different learning activities. In the ‘standard’ group, students continued to study and be tested over the entire word list. In the ‘repeated testing’ group, students continued to be tested over the whole list, but once a word pair was correctly recalled, it was dropped from further study. In the ‘repeated study’ group, participants repeatedly studied the entire list, but word pairs that were successfully recalled were not tested again. Finally, in the ‘drop’ group, students did not receive any further study or testing on words that were successfully recalled. By the beginning of the second phase, every student had successfully recalled each word pair in the list at least once, and thus performance during initial learning was equivalent across the four groups. However, when retention was measured after the second phase on a final test given 1 week later, a very different pattern of

results emerged: students in both the standard and repeated testing groups recalled approximately 80% of the word pairs, whereas students in the repeated study and drop groups recalled 36% and 33%, respectively. The critical difference between these groups is that the word pairs continued to be tested in both the standard and repeated testing groups, but not the repeated study and drop groups. These results indicate that repeated retrieval is key to promoting superior retention. Interestingly, the standard group did not demonstrate better retention than the repeated testing group, which suggests that continuing to study an item once it has been recalled has a negligible effect on retention.

One aspect of the Karpicke and Roediger¹⁶ experiment that deserves additional consideration is the use of multiple tests during learning. Repeated testing has been shown to promote better retention than taking a single test. For example, Wheeler and Roediger¹⁷ had students listen to a story while viewing pictures (60 in total) that corresponded with items mentioned in the story. Immediately afterwards, students were tested on the names of the pictures either three times, once or not at all. One week later, the group that took three tests recalled more of the names on a final test than the group that took one test, which in turn recalled more names than the group that was not tested at all. It is also important to note that in this study all the tests for the repeated testing group were administered consecutively, with no breaks between them. The benefits of repeated testing are even greater when tests are distributed over time.^{18,19} Indeed, the superiority of distributed or spaced practice over massed practice in promoting long-term retention, commonly referred to as the spacing effect, is highly replicable and quite robust.^{5,20} Thus, repeated testing that is spaced over time has great potential for improving learning in the classroom.²¹

Another important factor that influences the efficacy of testing in promoting retention is the format in which the test is given. Two broad classes of memory tests can be distinguished: recognition tests (e.g. multiple-choice, true/false, etc.), which involve selecting the correct response from a number of presented alternatives, and production tests (e.g. short-answer, fill-in-the-blank, essay, etc.), which require the test-taker to construct a response. Research has shown that production tests lead to better retention than recognition tests, presumably because production tests require more effortful retrieval of information from memory than recognition tests.^{3,22} For example, Butler and Roediger²³

manipulated the test format in an experiment that investigated the benefits of brief quizzes on long-term retention. Students viewed a series of three video lectures about art history on consecutive days. They engaged in a different activity after each lecture: taking a multiple-choice question (MCQ) test; taking a short-answer test, or studying a lecture summary. On a final test administered 1 month later, taking an initial short-answer test yielded better retention than either taking an initial MCQ test or studying the lecture summary. This study and others indicate that using production tests in the classroom will promote better retention than recognition tests.^{23,24}

One final factor that greatly improves the effectiveness of testing is feedback.^{25,26} Although testing improves retention in the absence of feedback, as shown in many of the studies described above, providing feedback enhances the benefits of testing by correcting errors and confirming correct responses.²⁷ Perhaps the most important aspect of feedback is the content of the feedback message. At the most basic level, feedback provides information about whether a response is correct or incorrect. However, much research has shown that simply providing information about the outcome of the response (i.e. right/wrong) does little to help students correct their errors, and that including the correct answer in the feedback message is critical.²⁸ Indeed, providing students with correct answer feedback can be essential to their learning from tests, especially when performance on an initial test is very low.²⁹ Another important aspect of feedback is the timing of the feedback message. Although there is some disagreement about whether feedback should be given immediately or after a delay (for review, see Kulik and Kulik³⁰), recent research shows that, at least in some cases, delayed feedback produces superior retention relative to immediate feedback.^{27,31}

Overall, research from the cognitive psychology laboratory can be used to make several recommendations about how to use testing to enhance learning in the classroom or other didactic settings. Firstly, educators should test frequently and repeatedly, perhaps using a brief quiz at the end of each learning session to space the tests out over time. Secondly, production tests, such as short-answer tests, should be used whenever possible instead of recognition tests, such as MCQ tests. Finally, feedback that includes information about the correct answers should be given after every test, but not necessarily immediately afterwards.

TESTING IN MEDICAL EDUCATION

Testing in medical education has largely served as an instrument of assessment. Tests are used to assign grades and to certify professional competence. Educators also use testing to evaluate the effectiveness of their curricula. In the ideal setting, tests serve as the culmination of the educational objectives that were outlined at the beginning of a course of study. Over recent decades, testing and assessment have evolved through various stages. The desire to measure the thought processes central to medical expertise has driven much of this change. The evolution of testing in medical education has been more thoroughly reviewed and analysed by van der Vleuten elsewhere.³² Here, we give a brief overview of concepts that are relevant to test-enhanced learning.

Changes in format and content have provided the dominant themes in the history of medical education assessment. The MCQ is one of the oldest and most commonly used test formats in medical education. It is easily and objectively graded and can be used to cover a wide array of information relatively quickly.³³ However, other test formats were developed because it was felt that MCQs do not accurately reflect the actual thought processes of the clinician.³⁴ The concern was that MCQs reflected only simple fact retrieval. Therefore, case-based formats with open-ended questions (OEQs), such as the modified essay question or other written simulations, were developed to measure the application of knowledge and rational problem solving.^{35–37}

A key assumption in this movement was that problem solving was a separate domain from factual knowledge.^{35,37} With further study, it became clear that the ability to solve problems was highly dependent on the knowledge on which the case was based – also known as case specificity.^{32,37} This finding caused a shift in focus to the memory constructs of medical knowledge that allow clinicians to access their knowledge base and solve problems.³⁵ The change in focus has led to new test formats such as key features problems, which emphasise those points that are absolutely necessary to solve a case, and script-concordance questionnaires, which assess the organisation of knowledge in novices relative to that in expert clinicians.^{38,39}

In the light of conclusions that expertise is based on knowledge and the organisation of that knowledge into usable networks, other issues related to test format were reconsidered. Studies of question format

showed high correlation between MCQs and OEQs.⁴⁰ Although there was an absolute difference in test performance, the relative correlation made each format predictive of the other. Therefore, the conclusion followed that response format is not as important as the content of the question.³⁵ Multiple-choice questions can be used just as well as OEQs to assess the foundations of expertise.³⁶

In addition to content and format, the educational context of assessment has been an important, although often understudied, theme in medical education testing research.^{32,41} The importance of this context is encapsulated in the oft-repeated aphorism that assessment drives learning. Progress tests offer one example of the potential educational impact of an assessment programme. The repeated testing of the curriculum allows students to measure their growth in knowledge and plan their study strategies in an ongoing manner.^{32,42} The test results also allow faculty members to gauge the effectiveness of the curriculum and its integration over time. Similar results have been found for in-training assessment programmes during postgraduate medical education.⁴³ Studies of educational impact provide a thorough summary of the indirect effects of testing and assessment, which include increased motivation to study, more efficient learning strategies, and improved accuracy in the measurement of curriculum efficacy.³²

APPLICATION OF TEST-ENHANCED LEARNING IN MEDICAL EDUCATION

If test-enhanced learning were to be implemented in medical education, the focus of testing and assessment might shift significantly. Tests would no longer be considered neutral tools of measurement, but rather active instruments to aid in the acquisition and retention of knowledge. Previous studies of the educational impact of testing, such as studies of progress tests or in-training assessments, have examined how testing measures the growth in knowledge of trainees or alters students' study strategies.^{32,42,43} As we discussed earlier, these results represent indirect effects of testing. Although these outcomes are positive and useful, they do not take into account the direct effects of testing. Perhaps one of the reasons that the mnemonic effects of testing have not been readily recognised in education involves the temporal separation of learning and testing activities. Assessment tests are typically given weeks to months after learning events, when the effects of tests to promote memory are minimised.¹¹ For tests to be used as memory-enhancement devices, they should be given relatively

soon after learning exercises have been carried out and should be derived specifically from the information learned.⁴ Test-enhanced learning binds testing directly to teaching and the educational process.

Studies of test-enhanced learning also draw conclusions about the influence of test format that differ from conclusions in the traditional medical assessment literature. The results of assessment tests show high correlations between MCQs and OEQs when they are both fact-based.⁴⁰ For assessment purposes, this relative relationship is adequate to judge a trainee's level of performance regardless of the test format. However, when testing is used to promote the retention of information, there is a clear contrast in the benefit of different question formats. Questions that require the learner to generate a response (i.e. fill-in-the-blank items or OEQs) produce better retention of information than questions that only require recognition (i.e. MCQs).²³ The theoretical reason for this effect seems to be driven by the amount of effort needed to recall the information.⁴ It remains to be seen whether multiple-choice formats that require more effortful retrieval, such as an extended matching test, might be comparable with production tests in promoting later retention.⁴⁴

The conclusions that can be drawn about the relative merits of various test formats depend on the relevant outcome measure. To judge the effectiveness of test-enhanced learning, an absolute increase in the amount of knowledge retained is the desired outcome. However, in the pure assessment approach, the relative correlation in performance between different formats allows rank ordering and pass/fail judgments to be valid whether MCQs or OEQs are used.

The medical education assessment literature has concluded that clinical expertise and problem solving rely heavily on a fund of clinical knowledge that is organised into a usable network.^{32,35} Test-enhanced learning seems particularly useful in that it facilitates improved retention of factual knowledge. Therefore, the use of tests to solidify this foundation of knowledge presents a promising tool with which to build clinical expertise. This technique may be particularly effective as students struggle to master complex and extensive sets of information, such as in physiology or pharmacology. Although this is applicable in medical education, we recognise that more than recall of information is needed to be an effective doctor. We need more studies that examine how tests affect the application of knowledge. One potential clinical example concerns evidence that teaching cardiac life support through simulation prevents the forgetting of this knowledge

over time.⁴⁵ This finding could be interpreted as a testing effect because the simulations serve as hands-on tests. The potential for testing to improve the organisation of knowledge into clinical scripts has also not been investigated. Future studies using instruments such as the script-concordance questionnaire will need to clarify how test-enhanced learning influences the formation of knowledge networks.

Despite the strong evidence from cognitive psychology laboratories, the theoretical possibilities of test-enhanced learning in medical education should be accepted cautiously. Several obstacles must be surmounted for test-enhanced learning to realise its promise in medical education. For example, the majority of evidence for the direct effects of testing has come from laboratory experiments. Of the more applied studies that have used standard educational materials, many have taken place in simulated classrooms.²³ However, at least one study showing the benefits of testing occurred in a university course.²⁴ Still, improved retention through testing has not yet been fully studied in many teaching situations, such as lectures, problem-based sessions or clinical activities. Identification of which educational opportunities lend themselves to test-enhanced learning is important because medical education uses a variety of settings and formats. We are currently engaged in research that aims to answer some of these questions.

Another issue that limits our ability to generalise the cognitive psychology studies of test-enhanced learning to medical education is the time-course of the testing in memory laboratories. Most studies of 'long-term' testing effects have used retention intervals (i.e. the time before the final examination) of less than 1 week, although a few have used longer intervals of up to 6 weeks.^{16,23,24} However, medical trainees need to retain information for months to years. Thus, the efficacy of testing in promoting retention over long periods of time has yet to be established.

Finally, test-enhanced learning faces the same challenge of all medical education interventions: how does it affect care of the patient? If doctors are better able to recall information on a test, does the same retrieval occur when needed during a patient encounter? Are students able to apply the information they recall? Testing techniques such as those using simulations and standardised patients may help to answer some of these questions as they approximate real patient care scenarios. However, if real patient outcome and care metrics were developed that could be followed before and after the testing

interventions, this would provide the strongest form of validation.

CONCLUSIONS

Test-enhanced learning refers to both the direct and indirect benefits of testing on long-term knowledge retention. Although laboratory experiments sometimes show dramatic positive effects of testing, research is just beginning to generalise these results to educational settings at all levels. However, we are optimistic that incorporating testing exercises of all sorts within the educational process can benefit learning, help create organised networks of knowledge, and promote the retrieval of information in contexts where it is needed. The types of test that seem to be most effective are those that involve active production of knowledge (rather than its more passive recognition), are spaced in time, and have feedback given at some delay after the test. Test-enhanced learning represents an educational intervention that is consistent with the current emphasis on using assessment to enhance educational practice in medical education. It could provide a welcome addition to the tools with which educators can help medical students, residents and practising doctors retain information and progress towards greater clinical expertise.

Contributors: all authors contributed to the concept and planing of the entire paper. HLR wrote the introductory section of the paper. ACB wrote the parts pertaining to testing research in cognitive psychology. DPL wrote the sections regarding testing in medical education and the application of test-enhanced learning in medical education. All authors extensively edited each portion of the paper. DPL collated the paper and performed the final edit prior to submission.

Acknowledgements: none.

Funding: grant support to HLR was provided by the James S McDonnell Foundation. This research was also supported by the Institute of Education Sciences, US Department of Education, through grant R305H060080-06 to Washington University in St Louis. The opinions expressed are those of the authors and do not represent the views of the Institute or the US Department of Education.

Conflicts of interest: none.

Ethical approval: not applicable.

REFERENCES

- 1 Wechsler Adult Intelligence Scale. Third Edition. San Antonio, TX: Pearson Education, Inc. 1997.

- 2 Roediger HL III, Karpicke JD. Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci* 2006;**17**:249–55.
- 3 McDaniel MA, Roediger HL III, McDermott KB. Generalizing test-enhanced learning from the laboratory to the classroom. *Psychon Bull Rev* 2007;**14**:200–6.
- 4 Roediger HL III, Karpicke JD. The power of testing memory: basic research and implications for educational practice. *Perspect Psychol Sci* 2006;**1**:181–210.
- 5 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**:354–80.
- 6 Dempster FN. Spacing effects and their implications for theory and practice. *Educ Psychol Rev* 1989;**1**:309–30.
- 7 Fitch ML, Drucker AJ, Norton AJ Jr. Frequent testing as a motivating factor in large lecture classes. *J Educ Psychol* 1951;**42**:1–20.
- 8 Leeming FC. The exam-a-day procedure improves performance in psychology classes. *Teach Psychol* 2002;**29**:210–2.
- 9 Gates AI. Recitation as a factor in memorizing. *Arch Psychol* 1917;**6**:1–104.
- 10 Jones HE. The effects of examination on the performance of learning. *Arch Psychol* 1923–24;**10**:1–70.
- 11 Spitzer HF. Studies in retention. *J Educ Psychol* 1939;**30**:641–56.
- 12 Bjork RA. Retrieval as a memory modifier: an interpretation of negative recency and related phenomena. In: Solso RL, ed. *Information Processing and Cognition: the Loyola Symposium*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. 1975;123–44.
- 13 Tulving E. The effects of presentation and recall of material in free recall learning. *J Verbal Learn Verbal Behav* 1967;**6**:175–84.
- 14 Glover JA. The ‘testing’ phenomenon: not gone but nearly forgotten. *J Educ Psychol* 1989;**81**:392–9.
- 15 Carrier M, Pashler H. The influence of retrieval on retention. *Mem Cogn* 1992;**20**:633–42.
- 16 Karpicke JD, Roediger HL III. The critical importance of retrieval for learning. *Science* 2008;**319**:966–8.
- 17 Wheeler MA, Roediger HL III. Disparate effects of repeated testing: reconciling Ballard’s (1913) and Bartlett’s (1932) results. *Psychol Sci* 1992;**3**:240–5.
- 18 Karpicke JD, Roediger HL III. Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances longterm retention. *J Exp Psychol Learn Mem Cogn* 2007;**33**:704–19.
- 19 Landauer TK, Bjork RA. Optimum rehearsal patterns and name learning. In: Gruneburg MM, Morris PE, Sykes RN, eds. *Practical Aspects of Memory*. London: Academic Press 1978;625–32.
- 20 Ebbinghaus H. *Memory: a Contribution to Experimental Psychology*. Ruger HA, Bussenius CE, transl. New York: Dover Publications, Inc. 1967. [Original work published 1885.] 1–33.
- 21 Dempster FN. The spacing effect: a case study in the failure to apply the results of psychological research. *Am Psychol* 1988;**43**:627–34.
- 22 Foos PW, Fisher RP. Using tests as learning opportunities. *J Educ Psychol* 1988;**80**:179–83.
- 23 Butler AC, Roediger HL III. Testing improves longterm retention in a simulated classroom setting. *Eur J Cogn Psychol* 2007;**19**:514–27.
- 24 McDaniel MA, Anderson JL, Derbish MH, Morrisette N. Testing the testing effect in the classroom. *Eur J Cogn Psychol* 2007;**19**:494–513.
- 25 Bangert-Drowns RL, Kulik CC, Kulik JA, Morgan M. The instructional effect of feedback in test-like events. *Rev Educ Res* 1991;**61**:213–38.
- 26 Kulhavy RW. Feedback in written instruction. *Rev Educ Res* 1977;**47**:211–32.
- 27 Butler AC, Roediger HL III. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Mem Cognit* 2008;**36**:604–16.
- 28 Pashler H, Cepeda NJ, Wixted JT, Rohrer D. When does feedback facilitate learning of words? *J Exp Psychol Learn Mem Cogn* 2005;**31**:3–8.
- 29 Kang SHK, McDermott KB, Roediger HL III. Test format and corrective feedback modulate the effect of testing on memory retention. *Eur J Cogn Psychol* 2007;**19**:528–58.
- 30 Kulik JA, Kulik CC. Timing of feedback and verbal learning. *Rev Educ Res* 1988;**58**:79–97.
- 31 Butler AC, Karpicke JD, Roediger HL III. The effect of type and timing of feedback on learning from multiple-choice tests. *J Exp Psychol App* 2007;**13**:273–81.
- 32 van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996;**1**:41–67.
- 33 Schuwirth LWT, van der Vleuten CPM. Different written assessment methods: what can be said about their strengths and weaknesses? *Med Educ* 2004;**38**:974–9.
- 34 Newble DI, Jaeger K. The effect of assessments and examinations on the learning of medical students. *Med Educ* 1983;**17**:165–71.
- 35 van der Vleuten CPM, Newble DI. How can we test clinical reasoning? *Lancet* 1995;**345**:1032–4.
- 36 Palmer EJ, Devitt PG. Assessment of higher-order cognitive skills in undergraduate education: modified essay or multiple-choice questions? Research paper. *BMC Med Educ* 2007;**7**:49.
- 37 Elstein AS, Shulman LS, Sprafka SA. *Medical Problem Solving: an Analysis of Clinical Reasoning*. Cambridge, MA: Harvard University Press 1978.
- 38 Page G, Bordage G, Allen T. Developing key-feature problems and examinations to assess clinical decision-making skills. *Acad Med* 1995;**70**:194–201.
- 39 Charlin B, Brailovsky C, Leduc C, Blouin D. The diagnostic script questionnaire: a new tool to assess a

- specific dimension of clinical competence. *Adv Health Sci Educ* 1998;**3**:51–8.
- 40 Norman GR, Smith EKM, Powles ACP, Rooney PJ, Henry NL, Dodd PE. Factors underlying performance on written tests of knowledge. *Med Educ* 1987;**21**:297–304.
- 41 Schuwirth LWT, van der Vleuten CPM. Changing education, changing assessment, changing research? *Med Educ* 2004;**38**:805–12.
- 42 Verhoeven BH, Verwijnen GM, Scherpbier AJJA, van der Vleuten CPM. Growth of medical knowledge. *Med Educ* 2002;**36**:711–7.
- 43 Ringsted C, Henriksen AH, Skaarup AM, van der Vleuten CPM. Educational impact of in-training assessment (ITA) in postgraduate medical education: a qualitative study of an ITA programme in actual practice. *Med Educ* 2004;**38**:767–77.
- 44 Case SM, Swanson DB. Extended-matching items: a practical alternative to free-response questions. *Teach Learn Med* 1993;**5** (2):107–15.
- 45 Wayne DB, Siddall VJ, Butter J, Fudala MJ, Wade LD, Feinglass J, McGaghie WC. A longitudinal study of internal medicine residents' retention of advanced cardiac life support skills. *Acad Med* 2006;**81** (10 Suppl):9–12.

Received 13 February 2008; editorial comments to authors 15 February 2008; accepted for publication 6 March 2008