Taking the sting out of assessment: is there a role for progress testing?

Debra Pugh[1] & Glenn Regehr[2]

CONTEXT It has long been understood that assessment is an important driver for learning. However, recently, there has been growing recognition that this powerful driving force of assessment has the potential to undermine curricular efforts. When the focus of assessment is to categorise learners into competent or not (i.e. assessment of learning), rather than being a tool to promote continuous learning (i.e. assessment for learning), there may be unintended consequences that ultimately hinder learning. In response, there has been a movement toward constructing assessment not only as a measurement problem, but also as an instructional design problem, and exploring more programmatic models of assessment across the curriculum. Progress testing is one form of assessment that has been introduced, in part, to attempt to address these concerns. However, in order for any assessment tool to be successful in promoting learning, careful consideration must be given to its implementation.

OBJECTIVE The purpose of this paper is to consider the implications of implementing progress testing within practice, and how this might promote or impede learning in the three phases of assessment (pre-test, pure-test and post-test).

METHODS We will examine the literature on how assessment drives learning and how this might apply to progress testing. We will also explore the distinction between assessment of learning and assessment for learning, including ways in which they overlap and differ. We end by discussing how the properties of an assessment tool can be harnessed to optimise learning.

CONCLUSIONS Progress tests are one potential solution to the problem of removing (or at least lessening) the sting associated with assessment. If implemented with careful thought and consideration, progress tests can be used to support the type of deep, meaningful and continuous learning that we are trying to instill in our learners.

Discuss ideas arising from the article at www.mededuc.com discuss

[1]Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada
[2]Faculty of Medicine, Centre for Health Education Scholarship, University of British Columbia, Vancouver, British Columbia, Canada

Correspondence: Debra Pugh, University of Ottawa, 501 Smyth Road, Box 209, Ottawa, ON, K1H 8L6, Canada. Tel: 613-737-8899. E-mail: dpugh@toh.on.ca

THE UNINTENDED CONSEQUENCES OF ASSESSMENT

In medical education, assessments of knowledge and clinical skills are frequently used to make judgements about competence. Although the goal of these assessments is to obtain an accurate estimate of ability, there are potential unintended consequences of assessment. For example, in an assessment of learning (AOL) model, the focus is on assigning grades or categorising learners into competent or not (i.e. pass or fail).[1] As such, there is a perceived component of social judgement or 'sting' associated with assessment from the learner's perspective.

As a result, the focus on AOL creates a tension for learners in that they may be oriented more toward efforts that result in higher test scores without necessarily promoting learning. The inherent pressures created by a testing environment can therefore lead to unwanted behaviours, which may range from adopting study habits that hinder deep understanding (e.g. cramming or memorising by rote) to outright cheating. Although the difference may seem subtle, when one approaches studying with the goal of passing a test rather than the goal of understanding the material, there can be a significant influence on learning. For example, studies have demonstrated that cramming is less effective at leading to long-term retention of material when compared with spaced learning.[2] The format of an assessment can also have an influence on learning, as students have been shown to alter their strategy for studying based on the format of an assessment, using more superficial approaches for tests thought to be designed to assess lower-order skills.[3,4]

In an AOL model, assessment is more likely to be treated by the learner as an obstacle that one must overcome. Thus, there is often little incentive for learners to revisit areas of weakness once the benchmark has been reached. As long as one is successful in meeting a predefined cut-score, regardless of whether one receives a very high score or only a marginally acceptable score, the message to learners is that those deficiencies are relatively insignificant and that they are ready to move on. Even if learners were motivated to improve, the feedback provided after a test is often not specific enough to guide learning.

Perhaps more concerning is the fact that several studies have demonstrated that cheating amongst medical students, including such behaviours as directly copying answers from another student and accessing unauthorised materials during a test, is widespread.[5-9] Even educators may be tempted to engage in dishonest behaviour to circumvent assessment processes, as evidenced by the recent conviction of 11 teachers in the USA on charges related to systematic cheating on standardised tests.[10]

Although assessment clearly plays an important role in ensuring the competence of learners, these subversive behaviours may lead one to question whether or not the current assessment culture is actually undermining the very learning that it is purportedly trying to promote. By contrast with the AOL model, an assessment for learning model (AFL) focuses on using assessment as an opportunity to provide feedback to learners to promote improvement.[11] Although not synonymous, AOL generally refers to 'summative' assessment, whereas AFL generally refers to 'formative' assessment. For the purposes of this paper, the term summative will refer to high-stakes assessments that emphasise achievement (e.g. grades), whereas the term formative will be used in the context of lower-stakes assessments that prioritise the provision of feedback. However, it is important to note that there is considerable overlap between these forms of assessment and both can be used in either AOL or AFL models.

An AFL model has the potential to mitigate some of the unintended consequences associated with an AOL model because of the shift from a focus on making judgements to a focus on creating opportunities for growth (i.e. emphasising excellence rather than minimal competence). In an AFL model, a programme of assessment can be designed with the goal of challenging every individual (whether high or low performing) to strive for continuous improvement. In this sense, tests become learning tools, hence the often-used term test-enhanced learning.[12]

THE ROLE OF PROGRESS TESTING

Changing the focus of assessment from AOL to AFL will require a programme of assessment that incorporates many tools in order to get a complete picture of learners' strengths and weaknesses for both formative and summative reasons.[1] One tool that appears to be garnering significant interest for use in an AFL model is the progress test. By design, the

blueprints of progress tests are based on a broad content domain (i.e. the complete domain of knowledge required for a programme). The tests are administered repeatedly to learners at different stages in their training in order to monitor their progress.[13] For example, some medical programmes administer a written progress test based on the entire undergraduate curriculum to the entire student body at regular intervals (e.g. four times per year).[14,15] Similarly, at least one residency programme offers, to all residents in the programme, an annual clinical skills progress test based on the objectives of training required for graduation.[16] A number of studies have demonstrated that progress tests can be used to chart growth of knowledge and clinical skills,[14–18] and allow for the provision of feedback.[19]

Progress tests appear to have tangible benefits beyond those of simply measuring progress. The use of written progress tests has been linked to improved performance in a national licensure examination,[20] whereas objective structured clinical examinations (OSCEs) used as progress tests have been shown to be useful in identifying residents at risk of subsequently failing a high-stakes national examination of clinical skills.[21]

Perhaps more importantly, however, the comprehensiveness of progress tests, coupled with the relatively low-stakes nature of the format, may serve to mitigate some of the subversive effects of AOL on learning. For example, because there is no functional limit on the content being assessed, strategies such as targeted studying, memorisation, cramming and teaching to the test are unlikely to be particularly useful. This appears to promote deeper learning strategies.[14,22] Additionally, meaningful feedback, which is often precluded in an AOL model due to test security issues, can be provided with limited risk as long as the item bank is sufficiently large. Thus progress tests have the potential to be important learning opportunities rather than hurdles to overcome. If learners can be persuaded to view progress tests in this way, then perhaps progress tests can be used to help take some of the 'sting' out of assessment.

In order for progress testing to support the type of learning intended by the curriculum, it is important to ensure that it is implemented effectively. To this end, the factors that modulate its effectiveness (i.e. by promoting or impeding learning) bear exploring.

HOW ASSESSMENT DRIVES LEARNING

The aphorism 'assessment drives learning' probably holds much truth. However, if one is to use assessment as a learning tool strategically, then one must consider how it promotes learning. The effects of assessment on learning can be divided into three phases: pre-test (i.e. tests incentivise learning), pure-test (i.e. tests directly lead to learning) and post-test (i.e. the feedback provided after a test leads to learning).[23] In the following section, each of these phases will be discussed in general terms, and in more specific terms as they might relate to progress testing.

Pre-test learning

The pre-test effects on learning, also termed the indirect effects of testing, refer to the extrinsic motivation provided by the anticipation of being tested. In other words, the imminent pressure caused by an impending test may provide learners with an incentive to study and attempt to learn the material to be tested in a consolidated way. In this sense, even summative testing may have a positive influence on learning by encouraging intentional efforts to learn material for the explicit purposes of retrieval. However, the way in which students approach learning may be negatively affected by a test if it promotes a superficial approach (e.g. memorising by rote rather than deeper understanding). For example, one of the often-cited benefits of progress testing is that it encourages continuous studying over cramming. This may be valuable because a review of over 180 papers comparing spaced versus massed practice supported the view that spaced learning is superior to massed practice.[24]

A useful theoretical model for the pre-test effects of summative assessment was developed by Cilliers et al.[25] This model postulates that there are two potential sources of impact that influence pre-assessment learning activities: task demands (e.g. task type and cues from lecturers) and system design (e.g. imminence of assessment and prevailing workload). These in turn influence cognitive processing and metacognitive regulation activities (e.g. perceived agency and interpersonal factors). For example, if an examination is imminent and students perceive it to be a high-stakes event, they may adopt lower-order cognitive strategies when studying, such as cramming or memorising by rote, even if that is not their usual preferred approach to learning, because

their goal is to maximise their chances of success. Studies about how the test format influences study behaviour would seem to support this. For example, students have been shown to use more superficial approaches to learning when they know they will be assessed using selected-response versus constructed-response formats[3] and part-task versus whole-task OSCEs.[4]

Although this framework was developed for summative assessment, one could speculate that, for progress testing, similar sources of impact (e.g. task type, assessment criteria, etc.) would theoretically influence approaches to learning. In support of this speculation, when McMaster's students' learning styles were studied using Mitchell's cognitive Behavior Survey, which attempts to assess learning style (memorisation versus concept learning), it was found that they continued to use deeper learning strategies rather than memorisation when progress tests were introduced.[14,26] Similarly, a study of performance on progress tests at Limburg University demonstrated that the use of a meaning-oriented study approach was positively correlated with progress test scores, whereas the use of memorisation or shallow processing was negatively correlated with scores.[22]

Of course, there is a risk that progress testing could actually undermine the value of the 'consolidating' effects of pre-assessment learning. When van Berkel et al.[22] compared study behaviours in relation to different test formats, students viewed block tests, but not progress tests, as an incentive to study (other than reviewing old tests), perhaps because it is so difficult to know how one would approach studying for a test of the entire curriculum, or perhaps because progress tests are not perceived to be important by learners. Moreover, although there is some evidence to suggest that spacing tests over time may encourage continuous study, this approach to learning may not always result in better test scores. In a study in which students were randomised to either a group that was assessed regularly (three times during a 10-week course) or a group that was assessed only at the end of the course, students who underwent more frequent assessments spaced their learning and devoted 69 hours more to overall study time than their peers, but they did no better on a final assessment.[27] In fact, students in the end-of-course assessment group reported spending significantly more time studying in the final week, which may challenge the long-held belief that spaced learning is preferable to cramming. However, although there was no demonstrated benefit of a spaced-learning approach, the study only examined the short-term effects, so it

is not clear if long-term retention would have been better in the spaced-learning group.

Pure-test learning

In addition to the effects of anticipating an upcoming test, the act of testing itself can lead to learning, something known as the testing effect. Several studies have demonstrated that tests can be used directly to influence learning above and beyond the indirect effects of studying.[28–31] This may be explained, in part, by the retrieval hypothesis, which suggests that the act of retrieving information leads to improved learning when compared with repeated study alone. It would seem that the act of retrieval helps to encode the information in a way that makes it more accessible for future retrieval.[32] A useful analogy to help understand this phenomenon is to consider books in a library. Libraries employ a classification system for shelving books in order to aid later retrieval. If, instead of using this system, books were simply placed on the shelves without consideration of how they would later be retrieved, it would be far more difficult to find them when needed. Similarly, when one practices retrieving information, it helps the brain to create a system for tagging or encoding it for future access.

The testing effect appears to not only enhance immediate retrieval, but also the retention of learned material for at least several months.[33,34] What is particularly interesting is that the testing effect seems to lead not only to better retention of facts, but also to improved application of knowledge, providing evidence that repeated retrieval may also lead to better understanding.[31,35,36]

There are important factors that seem to modulate the testing effect. For example, providing more opportunities for repeated testing of the same material has been shown to further enhance learning.[37] The testing effect also seems to be enhanced when the retrieval of knowledge is effortful; for example, by using constructed-response formats that require recall rather than simple recognition of information.[38,39] This may be explained, in part, by the theory of desirable difficulties,[32] which suggests that challenging tasks result in greater learning than simplistic tasks.

Progress tests would seem ideally suited for promoting the testing effect. Because progress test blueprints are based on an entire curriculum and are administered repeatedly, they expose learners to more content more often, which should enhance

the testing effect by prompting retrieval more frequently. Although progress tests typically employ a selected-response format (i.e. multiple-choice questions), other formats that require more effortful retrieval, including OSCEs, have been shown to be feasible.[40] There may be a case for increasing the use of constructed-response or OSCE formats to potentiate the testing effect through enhanced encoding. Although constructed-response formats tend to require more effort to mark, the benefits to learning may justify an increase in allotted resources.

Post-test learning

The final phase in which assessment can influence learning is after the test, when feedback is provided. Meaningful feedback is crucial in an AFL model because of its potential to influence metacognition. Feedback can help to highlight areas of strength and weakness, which may, in turn, be used by learners to direct their learning. Of course, in order to be meaningful the feedback must be based on sound data. If those data are in the form of test scores, then the test itself should have acceptable psychometric properties (e.g. scores that are reliable). Further, if the feedback is in the form of identifying areas of relative strength and weakness, then each area should have sufficient sampling to ensure reliability so that recommendations to focus on specific areas are based on a meaningful 'signal'.

Again, in this regard, progress tests are well positioned to be used as a source of both reliable data and rich feedback for learners. The results provided to students following progress tests are often accompanied by information about individual progress, summaries of areas of strengths and weaknesses,[19] and normative comparisons.[41] Because of their relatively low-stakes nature, learners can even be provided with copies of their progress tests to further guide their learning.[15] Of interest, progress tests also appear to be useful in predicting poor performance in national high-stakes examinations,[21,26] which can be used as an important source of feedback to promote learning.

Giving feedback, however, is not a straightforwardly simple process. For example, the type of feedback provided is important. Feedback focused on providing scores or data alone has not been shown to be sufficient to promote learning. In fact, providing grades, even when accompanied by narrative comments, may actually undermine learning. In a study in which grade-school children were provided with either grades alone, narrative comments alone, or both combined, only students who had received narrative comments alone showed a subsequent improvement when retested on similar tasks.[42] Further, it appears that the more detailed the feedback, the more likely students are to learn. When pre-medical students participated in a self-administered online multiple-choice test assessing biomedical knowledge, those who received detailed explanations about the correct and incorrect options outperformed those who only received information about the correct answer.[43] Although test security issues may preclude detailed feedback in some cases, there have been recent innovative efforts to provide increased feedback to learners, including in summative OSCEs.[44]

As an additional complication of the use of progress tests to provide feedback, it has been suggested that learners can be remarkably immune to feedback.[45,46] An important factor in accepting feedback appears to be the perceived credibility of the source providing feedback.[47] Telio et al.[48] proposed an educational alliance framework to explain this finding, in which they draw parallels between the feedback provider receiver relationship and that of a psychotherapist and patient. In this model, in order for feedback to be accepted and internalised by the learner, there needs to be a perceived alliance with the person providing the feedback. This alliance can help learners to allow themselves to be vulnerable enough to accept the guidance and criticism that is needed to help them achieve their potential. However, feedback can have an emotional impact, which can have detrimental effects on learning. Feedback that is focused on the task appears to have the most powerful influence on learning, whereas feedback that invokes the self (i.e. in the form of praise and punishment) appears to be less successful.[49] Perhaps this is because feedback that invokes the self carries with it social judgements that may threaten the educational alliance between learner and teacher, and may undermine the formative intentions of the assessment.

Summary

Many of the features of progress tests appear to align well with the properties that enhance desirable learning. Broad coverage of material makes cramming less feasible and may indirectly encourage more continuous and situated learning when compared with what occurs when one approaches a test with the goal of simply getting a good grade.

Repeatedly testing learners on the same material provides them with opportunities to practise retrieving information. Finally, progress tests allow for the provision of detailed feedback on performance. However, one must also consider the potential unintended consequences of introducing progress testing. It is possible that a progress test may prove to be a dis-incentive to studying if learners perceive it as unimportant or too overwhelming to prepare for. Although provision of feedback is possible, there are well-known complexities in ensuring that such feedback is incorporated meaningfully into further learning, so it shouldn't be taken for granted that the delivery of such feedback is by definition valued and incorporated.

This is not to say that progress tests are the only possible approach to addressing these issues. In fact, these principles could certainly be applied to other testing modalities, such as workplace-based assessments. By repeatedly offering experiences that challenge learners and allow them to make mistakes while providing them with help to improve, any form of assessment can be used to promote learning. Our point, rather, is that progress testing, if implemented carefully, is particularly well suited to a model of AFL.

HOW TO USE ASSESSMENT TO OPTIMISE LEARNING

Although the literature has drawn clear distinctions between AOL and AFL, it may be that there is more functional overlap between them than implied by the definitions. That is, it is likely that all assessment is used to assess a student's current state of learning to some degree, and all assessment has the potential to inform and support further learning. Importantly, there may also be discordance between the purported purpose of the test from the educator's perspective and how the learners interpret the test's purpose, so from the perspective of its influence on learner behaviour, the distinction between AOL and AFL may be in the eye of the beholder. In this sense, it is not a property of the test itself that determines whether it will be AOL or AFL, but rather the inferences that educators and learners make about the purpose of the test. The dichotomy, therefore, is somewhat artificial and for any given assessment situation it may be more useful to consider AOL and AFL as being relatively more in the foreground or background in the mind of the learner.

Thus, there is no question that different approaches to assessment may promote opportunities for learn-

ing more than others and, as such, progress tests can be designed to maximise their positive influence on learning. As discussed above, strategies that are likely to be effective in promoting learning when designing progress tests include: using constructed-response formats, offering frequent low-stakes testing opportunities, spacing tests and providing detailed feedback.

However, it is important to recognise that there are conceptual barriers to the use of AFL, such as the psychological consequences of knowing that one is being assessed. Even when assessments are labelled as being for learning (e.g. formative assessment), learners may view them as hurdles that they must overcome in order to succeed, rather than as learning opportunities.[50] To the learner, any assessment may be perceived as a high-stakes event where they are being judged. As such, the foreground of AOL may undermine the intentions of an AFL model. In order to move AOL to the background, the social meaning of assessment must be changed. In other words, learners need to be taught to view progress tests as opportunities to improve rather than as high-stakes events with potentially punitive consequences. We must create conditions of the assessment process that will ensure that we do not undermine our own efforts to foreground AFL in the minds of our students.

One way to accomplish this is to manage learners' expectations so that they come to expect regular feedback on their performance. However, again, the feedback that we provide will have an important influence on its perceived value as a learning opportunity. Thus we may consider withholding grades for progress tests (at least for a period of time) in favour of providing narrative feedback. This approach could help to minimise the implicitly summative nature of assessment, especially for medical trainees who have grown accustomed to being rewarded with good grades for a job well done. Learners could instead be encouraged to view their results from the progress tests as opportunities to explore potential areas of strengths and weaknesses and generate learning goals. Of course, learners may respond differently to assessment strategies – those who are performance-oriented may prefer summative assessments or benchmarks whereas those who are mastery-oriented may be more interested in formative assessments because they link learning with feedback and self-reflection.[51] However, if educators can set conditions that promote learners' orientation toward improving rather than proving themselves, then learners may be more

likely to embrace the learning opportunities provided by progress tests.

Practical issues may also act as barriers to the incorporation of an AFL model. Namely, educators facing significant time restraints may find it difficult to offer meaningful remediation or additional learning opportunities to address identified opportunities for learning. In an AOL model, resources tend to be reserved for learners who are identified as being in grave difficulty (e.g. those who have failed relatively high-stakes examinations), whereas in an AFL model, all learners are expected to receive guidance on how to improve. This added effort may stress over-worked physicians and educators. Focusing efforts on providing learners with the tools necessary to direct their own learning may help to offset some of this pressure, but we cannot ignore our responsibility as educators to guide this process if we are to most effectively support a positive learning response to the feedback that the progress test offers.

Finally, the issue of if and how to use progress tests and other formative assessments to make decisions about learners is a very practical concern. Although the primary purpose of these assessments may be formative, there is likely to be rich information that can be gained about learners.[52] Not all assessment can be exclusively *for* learning, as it is sometimes necessary to make summary judgements regarding learners' progress or continuation in a programme. However, using assessment data in this way may undermine attempts to implement an AFL model, as learners may shift their focus from learning to simply trying to overcome a hurdle. Despite educators' good intentions, it is possible that AFL does not really consider the perspective of the learner. It may be presumptuous to assume that learners will ever be willing to view assessment as a low-stakes, formative experience because of the implied judgements, or 'sting', inherent in all types of assessment. Instead, it may be useful to probe learners about how they use assessment to guide their learning and use these insights to modify programmes of assessment accordingly.[50]

CONCLUSIONS

As more and more educators embrace competency-based medical education (CBME), the AFL model is likely to continue to gain momentum.[53] This is not to suggest that there is not also a place for AOL. However, even when AOL is foregrounded (e.g. for

high-stakes assessment), the opportunity to promote continuous learning can and should be harnessed. For this to occur, it is necessary to find ways to take the sting out of the assessment experience; progress tests are offered as just one solution to the problem of removing (or at least lessening) this sting. Progress tests have many features that increase the likelihood that they will support the types of deep, meaningful and continuous learning that we are trying to instill in our learners. However, we must be careful that we do not undermine these efforts by assuming that progress testing will automatically fulfill this function. We must structure the tests intentionally to ensure that they shape learning practices through their pre-test, pure-test and post-test influences. We must ensure that the tests are supported by social constructions that foreground the AFL aspect of the assessments, and we must put in place the resources to ensure that students can take full advantage of the reflection and learning opportunities that the results of such tests offer. Unless we engage in these efforts, we run the risk that progress tests will devolve, in the minds of our students, to just another hoop they must jump through to survive our curricula.

REFERENCES

1 Schuwirth LWT, van der Vleuten CPM. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach* 2011a;**33** (6):478–85.
2 Pashler H, Rohrer D, Cepeda NJ, Carpenter SK. Enhancing learning and retarding forgetting: choices and consequences. *Psychon Bull Rev* 2007;**14** (2): 187–93.
3 Scouller K. The influence of assessment method on students' learning approaches: multiple choice

question examination versus assignment essay. *High Educ* 1998;**35**:453–72.

4 Lafleur A, Côté L, Leppink J. Influences of OSCE design on students' diagnostic reasoning. *Med Educ* 2015;**49** (2):203–14.

5 Dans PE. Self-reported cheating by students at one medical school. *Acad Med* 1996;**71** (1 Suppl):S70–2.

6 Baldwin DC, Daugherty SR, Rowley BD, Schwartz MD. Cheating in medical school: a survey of second-year students at 31 schools. *Acad Med* 1996;**71** (3):267–73.

7 Anderson RE, Obenshain SS. Cheating by students: findings, reflections, and remedies. *Acad Med* 1994;**69** (5):323–32.

8 Sierles F, Hendrickx I, Circle S. Cheating in medical school. *J Med Educ* 1980;**55** (2):124–5.

9 Kusnoor A, Falik R. Cheating in medical school: the unacknowledged ailment. *South Med J* 2013;**106** (8):479–83.

10 Bidwell A. Atlanta Scandal Could Be the Tip of a 'Test Cheating Iceberg'. U.S. News and World Report. 2015 April 2. http://www.usnews.com/news/articles/2015/04/02/11-former-atlanta-educators-convicted-in-standardized-test-cheating-scandal. [Accessed 22 September 2015.]

11 Schuwirth LWT, van der Vleuten CPM. General overview of the theories used in assessment: AMEE guide no. 57. *Med Teach* 2011b;**33** (10):783–97.

12 van der Vleuten CPM, Schuwirth LWT. Assessing professional competence: from methods to programmes. *Med Educ* 2005;**39** (3):309–17.

13 Freeman A, van der Vleuten C, Nouns Z, Ricketts C. Progress testing internationally. *Med Teach* 2010;**32** (6):451–5.

14 Blake JM, Norman GR, Keane DR, Mueller CB, Cunnington J, Didyk N. Introducing progress testing in McMaster University's problem-based medical curriculum: psychometric properties and effect on learning. *Acad Med* 1996;**71** (9):1002–7.

15 van der Vleuten CPM, Verwijnen GM, Wijnen HFW. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teach* 1996;**18** (2):103–9.

16 Pugh D, Touchie C, Humphrey-Murto S, Wood TJ. The OSCE progress test – measuring clinical skill development over residency training. *Med Teach* 2015;**38** (2):168–73.

17 Shen L. Progress testing for postgraduate medical education: a four-year experiment of American college of osteopathic surgeons resident examinations. *Adv Health Sci Educ Theory Pract* 2000;**5** (2):117–29.

18 De Champlain A, Cuddy M, Scoles P, Brown M, Swanson D, Holtzman K, Butler A. Progress testing in clinical science education: results of a pilot project between the national board of medical examiners and a US medical school. *Med Teach* 2010;**32** (6):503–8.

19 Coombes L, Ricketts C, Freeman A, Stratford J. Beyond assessment: feedback for individuals and institutions based on the progress test. *Med Teach* 2010;**32** (6):486–90.

20 Norman G, Neville A, Blake J, Mueller B. Assessment steers learning down the right road: impact of progress testing on licensing examination performance. *Med Teach* 2010;**32** (6):496–9.

21 Pugh D, Bhanji F, Cole G, Dupre J, Hatala R, Humphrey-Murto S, Touchie C, Wood TJ. Do OSCE progress test scores predict performance in a national high-stakes examination? *Med Educ* 2016;**50**(3):351–58.

22 van Berkel HJM, Nuy HJP, Geerligs T. The influence of progress tests on study behavior. *Instr Sci* 1995;**22** (4):317–33.

23 Dochy F, Segers M, Seger M, Gijbels D, Struyven K. Assessment engineering: breaking down barriers between teaching and learning, and assessment. In: Boud D, Falchikov N, eds. *Rethinking Assessment in Higher Education: Learning for the Longer Term.* Oxford: Routledge 2007;87–100.

24 Cepeda NJ, Pashler H, Vul E, Wixted JT, Rohrer D. Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychol Bull* 2006;**132** (3):354–80.

25 Cilliers F, Schuwirth LWT, Herman N, Adendorff H, van der Vleuten CPM. A model of the pre-assessment learning effects of summative assessment in medical education. *Adv Health Sci Educ Theory Pract* 2012;**17** (1):39–53.

26 Blake JM, Norman GR, Smith EKM. Report card from McMaster: student evaluation at a problem-based medical school. *Lancet* 1995;**345** (8954):899–902.

27 Kerdijk W, Cohen-Schotanus J, Mulder BF, Muntinghe FLH, Tio RA. Cumulative versus end-of-course assessment: effects on self-study time and test performance. *Med Educ* 2015;**49** (7):709–16.

28 Kromann CB, Jensen ML, Ringsted C. The effect of testing on skills learning. *Med Educ* 2009;**43** (1):21–7.

29 Larsen DP, Butler AC, Roediger HL III. Test-enhanced learning in medical education. *Med Educ* 2008;**42** (10):959–66.

30 Larsen DP, Butler AC, Roediger HL III. Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Med Educ* 2009;**43** (12):1174–81.

31 Larsen DP, Butler AC, Roediger HL III. Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Med Educ* 2013;**47** (7):674–82.

32 Bjork RA. Memory and metamemory considerations in the training of human beings. In: Metcalfe J, Shimamura AP, eds. *Metacognition: Knowing about Knowing.* Cambridge: The MIT Press 1994;185–205.

33 Rogers WA, Roediger HL, Agarwal PK, McDaniel MA, McDermott KB. Test-enhanced learning in the classroom: long-term improvements from quizzing. *J Exp Psychol Appl* 2011;**17** (4):382–95.

34 Kromann C, Bohnstedt C, Jensen M, Ringsted C. The testing effect on skills learning might last 6 months. *Adv Health Sci Educ Theory Pract* 2010;**15** (3):395–401.

35 Baghdady M, Carnahan H, Lam EWN, Woods NN. Test-enhanced learning and its effect on comprehension and diagnostic accuracy. *Med Educ* 2014;**48** (2):181–8.

36 Butler A. Repeated testing produces superior transfer of learning relative to repeated studying. *J Exp Pschol Learn Mem Cogn* 2010;**36** (5):1118–33.

37 Karpicke JD, Roediger HL. The critical importance of retrieval for learning. *Science* 2008;**319** (5865):966–8.

38 Kang SHK, McDermott KB, Roediger HL. Test format and corrective feedback modulate the effect of testing on memory retention. *Eur J Cogn Psychol* 2007;**19**:528–58.

39 Butler AC, Roediger HL. Testing improves long-term retention in a simulated classroom setting. *Eur J Cogn Psychol* 2007;**19**:514–27.

40 Pugh D, Touchie C, Wood TJ, Humphrey-Murto S. Progress testing: is there a role for the OSCE? *Med Educ* 2014;**48** (6):623–31.

41 McHarg J, Bradley P, Chamberlain S, Ricketts C, Searle J, McLachlan JC. Assessment of progress tests. *Med Educ* 2005;**39** (2):221–7.

42 Butler R. Enhancing and undermining intrinsic motivation; the effects of different feedback conditions on motivational perceptions, interest and performance. *J Educ Psychol* 1988;**58**:1–14.

43 Wojcikowski K, Kirk L. Immediate detailed feedback to test-enhanced learning: an effective online educational tool. *Med Teach* 2013;**35** (11): 915–9.

44 Harrison CJ, Molyneux AJ, Blackwell S, Wass VJ. How we give personalised audio feedback after summative OSCEs. *Med Teach* 2015;**37** (4):323–6.

45 Lockyer JM, Violato C, Fidler HM. What multisource feedback factors influence physician self-assessments? A five-year longitudinal study *Acad Med* 2007;**82** (10 Suppl):S77–80.

46 Eva KW, Munoz J, Hanson MD, Walsh A, Wakefield J. Which factors, personal or external, most influence student's generation of learning goals? *Acad Med* 2010;**85** (10 Suppl):S102–5.

47 Eva KW, Armson H, Holmboe E, Lockyer J, Loney E, Mann K, Sargeant J. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Adv Health Sci Educ Theory Pract* 2012;**17** (1):15–26.

48 Telio S, Ajjawi R, Regehr G. The educational alliance as a framework for reconceptualizing feedback in medical education. *Acad Med* 2015;**90** (5):609–14.

49 Hattie J, Timperley H. The power of feedback. *Rev Educ Res* 2007;**77** (1):81–112.

50 Heeneman S, Oudkerk Pool A, Schuwirth LWT, van der Vleuten CPM, Driessen EW. The impact of programmatic assessment on student learning: theory versus practice. *Med Educ* 2015;**49** (5):487–98.

51 Dijksterhuis MG, Schuwirth LW, Braat DD, Teunissen PW, Scheele F. A qualitative study on trainees' and supervisors' perceptions of assessment for learning in postgraduate medical education. *Med Teach* 2013;**35** (8):e1396–402 [Epub ahead of print].

52 van der Vleuten CPM, Schuwirth LWT, Driessen EW, Govaerts MJ, Heeneman S. 12 Tips for programmatic assessment. *Med Teach* 2014;**20**:1–6 [Epub ahead of print].

53 Holmboe E, Sherbino J, Long D, Swing S, Frank J. The role of assessment in competency-based medical education. *Med Teach* 2010;**32** (8):676–82.