Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: http://www.tandfonline.com/loi/imte20

Choosing and designing knowledge assessments: Experience at a new medical school

Adrian C. Freeman & Chris Ricketts

To cite this article: Adrian C. Freeman & Chris Ricketts (2010) Choosing and designing knowledge assessments: Experience at a new medical school, Medical Teacher, 32:7, 578-581, DOI: 10.3109/01421591003614858

To link to this article: https://doi.org/10.3109/01421591003614858

Published online: 23 Jul 2010.

2010; 32: 578–581 MEDICAL TEACHER

Choosing and designing knowledge assessments: Experience at a new medical school

ADRIAN C. FREEMAN & CHRIS RICKETTS
Peninsula College of Medicine and Dentistry, UK

Abstract

Background: Curriculum developers have a wide choice of assessment methods in all aspects of medical education including the specific area of medical knowledge. When selecting the appropriate tool, there is an increasing literature to provide a robust evidence base for developments or decisions.

Aim: As a new medical school, we wished to select the most appropriate method for knowledge assessment.

Methods: This article describes how a new medical school came to choose progress testing as its only method of summative assessment of undergraduate medical knowledge.

Results: The rationale, implementation, development and performance of the assessment are described. The position after the first cohort of students qualified is evaluated.

Conclusion: Progress testing has worked well in a new school. Opportunities for further study and development exist. It is to be hoped that our experiences and evidence will assist and inform others as they consider developments for their own schools.

The problem

As the number of medical schools increases and as existing schools revisit their educational practices, there is a need to review the best practice in assessment, decide what to adopt and consider how the best practice may need to be adapted to local circumstance and culture. Here, we describe how a new medical school developed its approach to assessing medical knowledge. This includes the decisions which needed to be made and some of the outcomes. The aim of this article is to provide curriculum and assessment developers with a rationale for choosing their own approach to assessing medical knowledge in the light of our experiences in a new medical school in the UK.

Our school had no traditions and no departments and was free to create a structure based on the best evidence available in medical education. What methods should we choose to assess the student's growth of knowledge?

Practice points

- In a medical degree course, knowledge assessments should provide a frequent feedback.
- It is feasible to have one consistent knowledge assessment method throughout the course.
- The most practical medical knowledge in the curriculum can be assessed by questions applied to clinical situations.
- Assessment material can be provided in-house and also in national and international question banks.
- Appropriate feedback can be provided whilst maintaining security of assessment material.

General principles

The undergraduate curriculum was designed around a problem-based learning model with a spiralling curriculum to allow review and repeated exposure to medical knowledge through the years. It also had a concept of early exposure to clinical content. Reviews of student progress were planned at the end of each year to see if students would cope with the increasing demands of each year. Separate strands of assessment were planned – medical knowledge; clinical skills; personal and professional growth and student selected modules. Our guiding principle was to embed the assessment process within the academic year as far as possible, rather than have high-stakes assessments at the end of the year: we articulated this as 'frequent look and rapid remediation'. We also considered the five aspects of assessment design propounded by Van der Vleuten of validity, reliability, educational impact, acceptability and cost (Van der Vleuten 1996).

What did we choose?

With these guiding principles, a review of the literature suggested that progress testing would be the best method to assess medical knowledge. The progress test is largely

Correspondence: A. C. Freeman, Peninsula College of Medicine and Dentistry, Peninsula Medical School, St Luke's Campus, Exeter EX1 2LU, UK. Tel: 44 139 226 2929; fax: 44 139 226 2926; email: adrain.freeman@pms.ac.uk

ISSN 0142–159X print/ISSN 1466–187X online/10/070578–4 © 2010 Informa Healthcare Ltd.
DOI: 10.3109/01421591003614858

independent of the curriculum. There can be no intensive pre-test revision strategies to distract it from the designed learning of the curriculum (Newble & Jaeger 1983). This prevents a superficial test-driven approach to study (Blake et al. 1996). There has been extensive use of progress tests internationally (Van der Vleuten 1996).

Data published from international sources demonstrate that:

(1) Medical students show a progress in scores over time compared with students from other faculties, i.e. the test measures learning in medicine.
(2) Medical students from a progress test university show a more consistent progress than students from a non-progress test university although both progress (Verhoeven et al. 1998).
(3) A similar pattern is demonstrated by comparison with medical students of traditional medical schools from different countries (Albano et al. 1996).

Progress testing sits with the assessment philosophy of frequent look and rapid remediation. It provides the students (cohort and individual) and staff with frequent measures of the effectiveness of the learning.

A number of institutions use progress tests, but the way they implement them varies in terms of frequency of testing, number of items in each test and the content domain. Practice varies from two to four tests a year, usually between 100 and 200 items per test, and the content domain may be either scientific knowledge or knowledge applied to particular clinical presentations.

We chose to use a progress test delivered four times a year because we wanted to sample, sufficiently to be able to respond to any changes in student learning, i.e. to have good educational impact. Each test contains 125 multiple-choice items per progress test and one right answer has to be chosen from the five options given. There is also a 'Don't Know' option and the negative marking for incorrect answers gave sufficient expected reliability. A typical single paper for a single year cohort produces a standard error mean (SEM) of less than 5%. Subsequent tests contain a different set of questions. The questions are in a clinical scenario format designed to assess application of medical knowledge including basic medical sciences.

The test is administered to all students. All students sit for the same test whatever be their year of study. Providing venues and materials to accomodate 900 students four times a year does have cost implications. However, this is the only knowledge test that the school uses. There are no separate block or subject tests to be devised or administered.

The knowledge level is set as that required by a newly qualified doctor. This means that students in their first year score low. As they advance through the course, scores increase indicating progress and growth of knowledge in the domain used by newly qualified doctors. Individual test scores are combined in an aggregation to give running grades of satisfactory, doubtful or unsatisfactory. It is these aggregate grades that are used for progression decisions at the end of the year.

What about norm referencing? Is there a cohort effect?

Data from published studies demonstrate a remarkable consistency in the performance in progress tests over the years and with different year entries (cohorts) (Blake et al. 1996; Van der Vleuten et al. 1996). Medical students seem to be a remarkably homogeneous group in terms of performance in knowledge tests.

The progress test is a multiple testing method. It is important to recognise that this is a longitudinal test. The students are judged on an aggregate grade over a series of tests and not just from a single test.

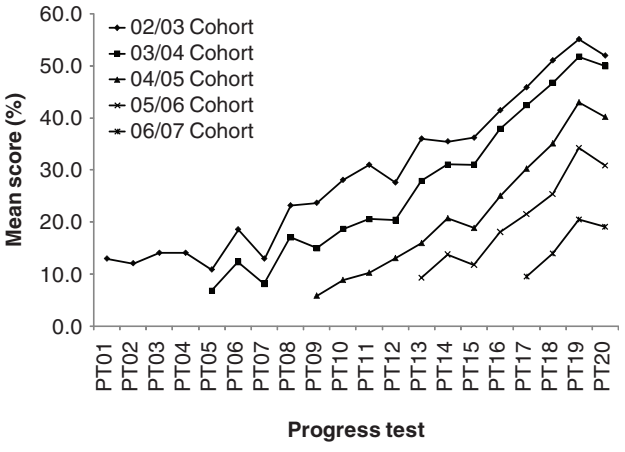
Why not criterion referencing?

Each test is made up of different questions and therefore has a different difficulty level. Attempts have been made to introduce criterion referencing. One study found that with norm referencing, the failure rate was reasonably constant but with absolute (criterion) referencing, the failure rate varied from 2% to 47% for different tests (Muijtjens et al. 1998). However, we were unwilling to use norm referencing for the high-stakes test in the final year (test acceptability) and hence we developed an approach to standard setting which is described elsewhere (Ricketts et al. 2009).

How do students perform over the years?

In the very early years, students have low scores because of their limited knowledge. As the years progress, their performance increases. The data from the progress tests allow defensible decisions to be made about student progress through the course. As a result of the frequent looks, there are no unanticipated surprises for students at year, or indeed course, end.

The tests have been effective since the start of the medical school course in 2002. The graph below shows the progress of the year cohorts as they joined the course. These graphs and their scores match other published data. The graph also illustrates the variation of the difficulty between tests. Scores are expressed as correct–incorrect and typical standard deviations for these cohort scores are about 7.5%.



The test material

All the questions for the test are structured in an applied knowledge setting. Each question starts with a clinical vignette with orientation and appropriate information, as per the best practice (Schuwirth et al. 1999). The aim is to test the application of knowledge and not just recalling simple facts. That application of knowledge is intended to mirror the activities of a practicing doctor, hence giving good expected validity.

The content of the test material is blueprinted against a modified version of a Professional and Linguistic Assessments Board (PLAB) blueprint (Tombleson et al. 2000). The PLAB blueprint is derived from a study of the working activity of junior doctors in the UK. This enhances the validity of the test. Monitoring of the test questions over time can indicate suitable coverage of the curriculum.

Question bank

Starting from nothing, we had to develop a bank of items. All items are stored in a secure commercial software program, which helps to generate the tests and also the machine marking of the optical mark answer forms. All items are reviewed by a Clinical Assessment Panel (CAP). This panel is a group of consultants and general practitioners (GPs) who regularly work with newly qualified doctors. By creating a large enough group (20) of clinicians, it allows flexibility to attend to service delivery as well as these medical school activities. CAP meets once a month and assures quality for each question before it enters the bank. Again this enhances the validity of the test as these are the clinicians who are working with doctors in training. They are then the best judges of the appropriateness of the question material, difficulty of the item, etc. Apart from reviewing new questions, CAP is also tasked to review the items post-test that have not been performed well, as indicated by the analysis of question performance.

New items are generated mostly in item writing workshops attended by CAP members, teaching staff, etc. (Case et al. 2001). We have been part of the International Database for Enhanced Assessments and Learning (IDEAL), an international consortium of medical schools which gave access to a large question bank which was essential in the early stages. Although the questions from that bank were not directly transferable, it provided the core material which we could manipulate into the format required for our tests (IDEAL 2006).

Additionally, we became fully involved with the Universities Medical Assessment Partnership (UMAP) consortium as it developed. This is a large grouping of UK medical schools developing and sharing assessment material (UMAP 2009). We are hopeful that using this material will allow some indirect comparison of standards between our school and other UK schools.

Feedback to students

As the questions are in a secure bank and we share (some of) the assessment material with other consortia, we cannot release the details of the questions to the students. For each question, we produce a 'learning point', which indicates the area of knowledge the question was testing. After each test, the student receives a comprehensive report electronically. This includes the student's scores in the test; their grades for the test; their consequent running aggregate grades; the comparison of performance to the year cohort and a list of learning points for each question and whether they got it right, wrong or did not answer.

Twice a year, each student meets an academic tutor designated to review his progress.

Data collection

Such a test system allows for considerable data collection. As a consequence of blueprinting, we can link the test performance with curriculum activity and provide a detailed feedback to the teaching staff. Each test is analysed to check for item and test performances including facility and point biserial scores for every question; any identifiable bias between sub-groups of students is looked for.

Quality assurance

A panel consisting of academic, clinical and administrative staff oversees the whole process. The panel reviews each test before and after it is delivered. All panel activities are recorded and are available to the school management committee. A dedicated external examiner reviews the progress test and has access to all data.

Conclusions

After our first students graduated, we could see that the progress test conducted in our school was in the same manner as other schools from the published literature. Progress testing is a universal method applicable to different countries and inevitably different curricula. This assessment does encourage appropriate learning, and a study conducted in our school confirmed that students with a deep approach to learning perform better in these tests than students with a superficial approach (Mattick et al. 2004). We would hope that there would be opportunities in the UK to develop cross-institutional tests as has been achieved in the Netherlands (Muijtjens et al. 2008). Our early work with UMAP will hopefully encourage such developments. We are looking at the effectiveness of sharing items in more detail and thereby reducing costs as well as comparing standards. We are carrying out further work on the consequences of remediation offered by this testing and the enhanced opportunities for feedback. Information technology is advancing and there should be scope for on-line delivery of progress tests, but there are significant implications for summative assessments which have to be taken into account.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Notes on contributors

ADRIAN C. FREEMAN is a Lecturer in assessment at the Peninsula College of Medicine and Dentistry. His clinical speciality is family medicine and he is an examiner for the Royal College of General Practitioners.

CHRIS RICKETTS is the Director of assessment at the Peninsula College of Medicine and Dentistry and a National Teaching Fellow. He has worked on assessment for over 20 years and has particular interest in computer-assisted assessment, generalisability analysis and standard setting.

References

Albano M, Cavallo F, Hoogenboom R, Magni F, Majoor G, Manenti F, Schuwirth L, Stiegler I, Van der Vleuten CPM. 1996. An international comparison of knowledge levels of medical students: The Maastricht progress test. Med Educ 30:239–245.

Blake JM, Norman GR, Keane DR, Mueller CB, Cunnington J, Didyk N. 1996. Introducing progress testing in McMaster University's problem-based medical curriculum: Psychometric properties and effect on learning. Acad Med 71(9):1002–1007.

Case SM, Holtzman K, Ripkey D. 2001. Developing an item pool for CBT: A practical comparison of three models of item writing. Acad Med 76(10):111S–113.

IDEAL-HK. [Published 2006]. Available from: http://www2.hkwebmed.org/webmed/home1.htm

Mattick K, Dennis I, Bligh J. 2004. Approaches to learning and studying in medical students: Validation of a revised inventory and its relation to student characteristics and performance. Med Educ 38(5):535–543.

Muijtjens A, Hoogenboom R, Verwijnen GM, Van der Vleuten CPM. 1998. Relative or absolute standards in assessing medical knowledge using progress tests. Adv Health Sci Educ Theory Pract 3:81–87.

Muijtjens A, Schuwirth L, Cohen-Schotanus J, Van der Vleuten C. 2008. Differences in knowledge development exposed by multi-curricular progress test data. Adv Health Sci Educ 13(5):593–605.

Newble DI, Jaeger K. 1983. The effect of assessments and examinations on the learning of medical students. Med Educ 17(3):165–171.

Ricketts C, Freeman A, Coombes L. 2009. Standard setting for progress tests: Combining external and internal standards. Med Educ 43(6):589–593.

Schuwirth LWT, Blackmore DE, Mom E, Van Den Wildenberg F, Stoffers HEJH, Van Der Vleuten CPM. 1999. How to write short cases for assessing problem-solving skills. Med Teach 21(2):144–150.

Tombleson P, Fox R, Dacre J. 2000. Defining the content for the objective structured clinical examination component of the Professional and Linguistic Assesments Board examination: Development of a blueprint. Med Educ 34:566–572.

UMAP. (Published 2009). Available from: http://www.umap.org.uk/

Van der Vleuten C. 1996. The assessment of professional competence: Developments, research and practical implications. Adv Health Sci Educ 1:41–67.

Van der Vleuten C, Verwijnen G, Wijnen H. 1996. Fifteen years of experience with progress testing in a problem-based learning curriculum. Med Teach 18:103–109.

Verhoeven BH, Verwijnen GM, Scherpbier AJJA, Holdrinet RSG, Oeseburg B, Bulte JA, Van der Vleuten CPM. 1998. An analysis of progress test results of PBL and non-PBL students. Med Teach 20(4):310–316.