Techniques in medical education

Recent developments in assessing medical students

Susan L Fowell, John G Bligh

Summary

Most medical schools in the UK are revising their undergraduate courses in response to the recommendations published by the General Medical Council Education Committee in *Tomorrow's doctors*. However, achievement of the objectives of curricular change is attendant on revision of the assessment process. This paper reviews traditional and more recently developed methods for assessment of medical education in the light of the General Medical Council's recommendations which relate specifically to summative assessment of the core curriculum. The importance of reliability and validity is highlighted, and the case for criterion-referenced assessment is examined.

Keywords: education, undergraduates, assessment

There is currently considerable interest and discussion about assessment in medical education at both the undergraduate and postgraduate level. (Literature searches on Medline identified over 200 papers related to assessment in medical education published between 1994 and 1996.) In the UK, increases in student numbers, diversification of student intake and the introduction of National Vocational Qualifications (NVQs) have combined to raise the profile of assessment in higher education in general.¹ Although it may be argued that these factors impinge less directly on medical schools than on other departments in academic institutions, the Employment Department's report *Assessment issues in higher education*¹ sets out recommendations at the institutional, faculty and individual level which will inevitably change the teaching climate within universities and, as a consequence, will affect medical faculties. In addition, postgraduate bodies and Royal Colleges are reviewing their assessment procedures, for example, the membership exam of the Royal College of General Practitioners (the MRCGP exam) has undergone extensive review in response to specific concerns about assessment.² However, the main driving force for review of assessment methods in medical education is undoubtedly the 1993 report by the General Medical Council (GMC) Education Committee *Tomorrow's doctors*.³ In summary, the report recommends an undergraduate curriculum comprising a common core which will be followed by all students, with optional 'special study modules' to allow flexibility between medical schools and specialisation by individual students.⁴ The implications of these recommendations have been discussed elsewhere⁵ and most medical schools in the UK are revising their curricula in response.

It has been established that the mode of assessment influences the learning style of students (assessment drives learning),^{6,7} and it has been shown that medical students are susceptible to this influence.⁸ Medical schools are therefore seeking ways to change their assessment procedures.

UK medical schools are not alone in wishing to change assessment. The National Medical Licensing Boards of Canada and the USA are continually reviewing their assessment procedures and there have been several interesting developments in recent years, for example, extended matching items questions, and key features questions, which will be considered later in this review, and the use of standardised patients.^{9–11} In addition, innovative medical schools in North America and Europe have invested considerable effort into devising assessment strategies which support their educational philosophies.^{12–15}

The aim of this review is to consider recent developments in assessment methods in relation to the relevant recommendations of the GMC. The theory of assessment is well covered elsewhere,⁷ and a review of assessment in general is beyond the scope of this article. Assessment of medical students has been reviewed previously,^{16,17} and either reference provides an excellent introduction for any reader who is new to this field.

When designing assessments, it is important to determine the purpose of assessment before selecting methods. Put more simply, one should decide why learners are to be assessed and what they should be assessed on, before deciding how they will be assessed. As a starting point we will focus on the **summative** assessment (box 1) of the core curriculum and address the three categories of objectives identified by the GMC; namely, knowledge, skills and attitudes. The selection of suitable assessment methods or subsequent evaluation of assessments for any of these aspects should be guided by the **reliability** and **validity** of the assessments (box 1).

Assessment of core knowledge: moving beyond factual recall

The major principles influencing the choice of assessment methods for core knowledge are reducing the factual overload on students, assessing knowledge at

Department of Health Care Education,
The University of Liverpool, 3rd Floor
University Clinical Department,
Duncan Building, Liverpool
L69 3GA, UK
SL Fowell
JG Bligh

Accepted 6 May 1997

Glossary of assessment-related terms

- **formative assessment:** assessment intended to provide feedback to students about their progress; formative assessment does not affect their progress through the course
- **summative assessment:** assessment designed to measure what the students have learned on the course; summative assessment usually occurs at the end of a module, phase or stage of a course and acts as a barrier to progress. Summative assessments must be passed before students can proceed to the next stage or be deemed to have passed the course
- **reliability:** a measure of whether the assessment (or test) is consistent and accurate; examines the extent to which factors such as examiners, questions, occasions affect the marks (or scores) awarded
- **validity:** a measure of the extent to which the test actually measures what it is intended to measure. For example, the question 'Discuss the roles of insulin and glucagon in glucose homeostasis' is not a valid assessment of a candidate's ability to manage diabetes
- **face validity:** the acceptability of the assessment to the examiners and candidates, ie, does the assessment appear relevant, is the wording appropriate?

Box 1

Setting pass-marks

- **norm-referenced:** candidates' scores are ranked in order (usually a normal distribution). The pass mark is adjusted so that a set percentage of students fail, eg, 10% of candidates. The pass mark varies on each application of the test, but the pass rate stays the same
- **criterion-referenced:** the pass mark (or cutting score) is set to an absolute standard (which is usually pre-determined). The pass mark stays the same on each application of the test, but the pass rate varies

Box 2

higher levels than simple factual recall, and rigorously assessing the core. When examining core knowledge, assessment procedures must ensure that all students meet a minimum standard, which ultimately relates to their future role as pre-registration house officers. This raises the issue of **standard-setting** or determination of the pass mark (box 2).

The advantages of **criterion-referenced**, as opposed to **norm-referenced**, standard setting have been discussed previously¹⁷ and are reiterated in the figure, which illustrates how using a norm-referenced pass-mark can allow students who do not meet minimum standards to 'pass' an assessment. In addition, the use of the traditional pass mark of 40–50% seems incompatible with rigorous assessment of core, as in effect, students can pass with knowledge of only half of the core content.

Finally, the recommendations state explicitly that assessment should test acquisition and utilisation of core knowledge. The core knowledge objectives must thus be tested at cognitive levels beyond mere recall. Bloom's taxonomy of learning objectives¹⁸ may provide a useful framework here. The challenge in assessing core knowledge is to develop objective methods which test higher cognitive skills across the whole of the curriculum.

Assessing knowledge: traditional and innovative methods

ESSAYS

Methods of assessment have traditionally included essays and multiple-choice questions. The considerable disadvantages of essays for assessment are discussed elsewhere.^{19–20} In short, essay questions are easy to set, difficult to mark and have low reliability. Many readers may have taken and set final examinations which were assessed primarily by one or more essay papers, in which candidates typically choose, for example, four questions from a total of six. Such exams encourage strategic learning, question-spotting and lead to undesirable learning patterns, and are incompatible with the requirements for rigorous assessment of core stated above.

MULTIPLE-CHOICE QUESTIONS

There has been, and will continue to be, much debate about the use of multiple-choice questions (MCQs) for the assessment of medical students.^{21–22} In their favour, MCQs are reliable, are easy to mark and can be used to sample a large part of the curriculum. Although a frequent criticism of the use of MCQs is that they mainly test factual recall, this is not a fault of MCQs *per se* but a reflection of the way in which such questions are often used. It is possible to devise MCQs which test higher cognitive skills and evidence exists that these skills are utilised in answering such questions.^{23–26} MCQs appear easy to write, but devising good MCQs takes considerable effort. Several references contain useful guidelines for writing MCQs and examples of their use.^{27–28} Use of well-written MCQs seems most appropriate for assessing core knowledge as this will allow a large number of the core objectives to be assessed at different cognitive levels.

TESTING HIGHER COGNITIVE LEVELS OF KNOWLEDGE—SHORT ANSWER QUESTIONS AND MODIFIED ESSAY QUESTIONS

Another criticism of using MCQs to assess medical students, which cannot be denied by even their most enthusiastic advocates, is that the choice of a small number of alternatives may provide clues to the answer which the candidates would not have generated if left to their own devices (cueing).²⁹ As patients do not usually present with a list of five alternative treatments or diagnoses and, in real life, clinicians are required to generate their own options, cueing reduces the validity of MCQs for assessing the application of knowledge in areas such as diagnosis and treatment. The use of short answer questions³⁰ provides one approach if cueing is a concern. Short answer questions can be considered to fall in between MCQs and essays - they sample a large part of the curriculum but are less reliable and less easy to mark than MCQ. Similar considerations apply to modified essay questions(box 3).³¹

EXTENDED MATCHING ITEMS QUESTIONS

An alternative approach to overcome the problem of cueing is to increase the number of options. This requires the candidate to generate the answer to the question, as it is not feasible to consider each option in turn in the time allowed. Extended matching items questions were developed by Case and Swanson²⁹ and are now used in the US Medical Licensing Examination. This format represents a compromise between free-response questions (eg, essays) and MCQs, and offers an objective assessment which is reliable and easily marked.

Figure Comparison of norm- and criterion-referenced pass marks. These diagrams illustrate how norm-referencing may allow students who do not meet minimum standards to 'pass' and compare an absolute criterion-referenced pass mark with a norm-referenced or relative pass mark. In a year where students perform well, some students who are above the minimum standard will fail a norm-referenced assessment. More worryingly, if students perform badly, norm-referencing will allow students who do not meet the minimum criterion to pass

Example of typical modified essay questions*

Jane Smith presents her 3-year-old son Simon for the fourth time in six months complaining that he has diarrhoea. She says she is tired all the time, bursts into tears and wants you to refer Simon to a specialist.

Describe your reply, and your management of Jane and Simon

*Based on format used for MRCGP Examination. Candidates required to answer 12 questions in 2 h

Box 3

Extended matching items questions consist of four elements; the *theme*, *options*, *lead in* and *stem* (see box 4). Questions relating to a particular topic or complaint are arranged in *themes*, eg, abdominal pain. To avoid cueing, more than five and up to 26 options are presented for each theme. (In practice, the number of options is limited by the number of alternatives available on optical marking sheets.) The lead-in sets the scene for the questions and the stems are usually case vignettes. Extended matching items questions are considered to be a reliable and potentially valid form of assessment of higher levels of knowledge.²⁹

KEY FEATURES QUESTIONS

Key features questions were developed by Page and Bordage to replace patient management problems³² for assessing clinical decision-making skills in the Common Qualifying Examination of the Medical Council of Canada. This development was prompted by studies of the psychometric properties of patient management problems and research into clinical problem solving, which showed that patient management problems have low reliability and that problem-solving skills are specific to a particular problem and are not transferable, ie, these skills are case or **content specific**.³³

Key features questions were based on the premise that clinical decision-making skills depend on recognising a few elements or **key features** of the problem.³⁴ Guidelines for writing and developing key features questions are provided elsewhere.^{33 35} In brief, the process requires generating the key features of a problem (box 5) and devising questions to elicit these responses (box 6). These questions can include both short answer questions and extended options as appropriate. Studies of the use of key features questions indicate high reliability³⁶ and these questions were incorporated in the Medical Council of Canada's Qualifying Examination in 1992.

Assessment of core skills by objective structured examinations

A key factor to consider in the assessment of clinical skills, including communication skills, is the requirement for rigorous assessment of core. Assessment must thus be wide ranging and cover a variety of tasks. There is an extensive literature on the assessment of clinical skills,^{37 38} and many assessments are variants of the objective structured clinical examination.

OBJECTIVE STRUCTURED CLINICAL EXAMINATIONS

Objective structured clinical examinations were first developed in 1979^{39 40} and there is an wealth of literature relating to their application both in medical specialties^{41 42} and fields other than medicine.^{43 44} They are considered to have high reliability and validity for assessment of clinical skills⁴⁵⁻⁴⁷ and guidelines for their implementation have been published.^{48 49} Their psychometric properties have been studied extensively and aspects such as station length⁵⁰ and patient and student gender on performance⁵¹ have been examined. The value of providing feedback on performance has been discussed,^{52 53} and the application of different question formats has been examined.⁵⁴

COMMUNICATION SKILLS

Communication skills are seen as an essential component of the core curriculum. These skills can be included as part of objective structured clinical examination stations; however, it has been suggested that it is more valid and reliable to use stations specifically designed to assess communication, rather than considering communication as part of an 'add on' to assessment of other skills such as history taking.⁵⁵ The use of specific stations or objective structured behavioural examination has been described previously.^{55 56} A more recent development in the assessment of communication skills is the objective structured video examination which examines the ability to recognise types of communication being utilised in a particular situation, to identify the consequences of the approach used, and to suggest appropriate alternative strategies. Each student is shown a video-recorded interaction between a clinician and patient, and then answers written questions that relate to the interaction.⁵⁶

Assessment beyond specific clinical skills: clinical competence

Although objective structured clinical examinations are well accepted as methods of assessing clinical skills, some concern has been voiced that the short station format does not offer an opportunity to assess history taking and physical examination, abilities which many clinicians consider to be key components of clinical competence.⁵⁷ There is much discussion about assessment of clinical com-

Extended matching items questions**Theme:* investigation of abdominal pain*Options:*

- CT scan
- liver function tests
- ERCP
- plain abdominal X-ray
- erect chest X-ray
- serum amylase
- HIDA scan
- ultrasound scan of liver and biliary tree

Lead-in: For the patients described below, which is likely to be the SINGLE most helpful investigation? *Each answer may be used once, more than once, or not at all*

Stems:

- a 45-year-old woman is admitted with colicky right upper quadrant pain radiating to the back and shoulder tip. She is tender with guarding in the right upper quadrant. She has a pyrexia of 37.8°C and an elevated white blood cell count of $18 \times 10^9/l$
- a 56-year-old woman is awaiting a cholecystectomy for chronic cholecystitis. She returns to the clinic with a three-week history of painless jaundice, dark urine and pale stools
- a 52-year-old man is admitted with a 36-h history of colicky abdominal pain, vomiting faeculent fluid and absolute constipation. Examination reveals a distended abdomen with tinkling bowel sounds
- a 52-year-old woman is admitted following the sudden onset of severe epigastric pain, radiating to the back. She is slightly more comfortable sitting forwards. She has been vomiting profusely. On examination she is tender with guarding throughout the upper abdomen and bowel sounds are absent. She is known to have gallstones but has never had pain like this.

*Taken from University of Liverpool fourth year examination in Medicine and Surgery, 1996

Box 4**Example of key features problem**

For a woman of childbearing age presenting with history of pain in the right iliac fossa, the candidate should:

- generate as leading diagnoses: acute appendicitis, urinary tract infection, ectopic pregnancy
- be able to identify important features in the history and examination which distinguish between these diagnoses

Box 5

petence in medical education.^{37 38 58 59} This section provides a broad overview of methods for assessing clinical competence, considering the traditional methods used (orals and global ratings), and highlighting recent developments which are appropriate for assessing competence.

ORAL EXAMINATIONS

No consideration of assessment of clinical competence would be complete without mentioning oral examinations and vivas, if only to restate their disadvantages. The application of oral examinations in assessment in medical education has been reviewed,^{60 61} and it has been demonstrated that orals have low reliability as assessments of clinical competence. To be precise, when the ratings or marks awarded to candidates by different examiners are compared there is low reliability between the ratings. It is thought that this is, in part, a consequence of low reliability between examiners (inter-rater reliability); some examiners tend to mark generously (doves) and some have a tendency to award low marks (hawks). There is also evidence that the low reliability is due to content specificity, reflecting the fact that competence is related to the case or topic covered in the oral.

Concerning the validity of oral ratings, studies indicate that the mark awarded to a candidate may reflect factors other than the candidate's clinical competence; namely anxiety, percentage of words contributed to the discussion by the candidate, the examiner's visual impressions of the candidate or the candidate's self-confidence.⁶²⁻⁶⁵ Rating is also prone to the halo effect in which, if a candidate performs well (or badly) on one assessment criterion, this influences the rating given on subsequent criteria which are judged higher (or lower) than the actual performance. These poor psychometric properties led the US National Board of Medical Examiners to drop orals from Part II of the licensing examination in 1963.⁶⁰

Despite these drawbacks, orals are still widely used in the UK by a number of medical schools and Royal Colleges. To some extent the oral is seen as a necessary 'rite of passage'. Wakeford *et al* recently reviewed the use of orals as part of the MRCGP exam and they provide a useful set of guidelines for improving orals.⁶⁶ This should be required reading for all examiners involved in oral or *viva voce* examinations.

GLOBAL RATINGS

Global ratings describes any summative judgement of a student's performance completed by a supervisor after a period of contact, for example, during a clinical attachment or clinical firm. These usually involve completion of a form containing rating scales for each aspect of clinical performance.

The use of global ratings has been discussed and reviewed extensively^{34 67 68}; they are generally considered to be unreliable as an assessment method. In particular, they are susceptible to the halo effect and often reflect the degree to which the supervisor and student get on during the period of contact. Global ratings are, however, widely used as an assessment method and have some attractive features. Guidelines for increasing their reliability have been described,^{67 68} and should be consulted by any examiner intending to include global ratings in assessment.

OBSERVED LONG CASES AND THE OBJECTIVE STRUCTURED LONG EXAMINATION RECORD

An observed long case is a development of the traditional oral case presentation, first developed at the University of Adelaide to replace oral examination of final year medical students.⁶⁹ The objective structured long examination record is an attempt to structure and improve the reliability of the traditional 'long case'.⁵⁷ The examiner completes a 10-part record throughout the observation. This record covers aspects such as presentation of the history, physical examination, investigation and management. This approach has much to recommend it as an additional assessment of clinical competence in physical examination and history taking, offering a compromise between the reliability of the multi-station objective structured examination record and the face validity of the extended patient-student interaction of the long case.

SIMULATED SURGERIES

Simulated surgeries are a development of the objective structured clinical examinations which have been proposed and evaluated as an assessment of clinical competence in the MRCGP.^{70 71} In a simulated surgery, the candidate is observed by one, or ideally two, examiner(s) whilst consulting with a series of standardised patients in a realistic general practice setting. This differs from the traditional objective structured clinical examination in that it is the standardised patients who visit 'stations' in rotation, whilst the candidate stays in one place.

Example question to elicit these key features

Case

Mary, a 25-year-old woman, presents with a 12-h history of pain in her right iliac fossa

- Which three diagnoses would you consider?
- With respect to these diagnoses, select three elements of her history which you would want to elicit from the following list: anorexia, bowel disturbance, contraceptive history, date of last menstrual period, dysuria, family history, frequency of micturition, haematuria, headache, history of urinary tract infections, onset of pain, polydipsia, recent weight loss, shift of pain, sore throat, vomiting
- List two features you would look for on physical examination

Box 6

Simulated surgeries can potentially assess a range of skills, including physical examination, doctor–patient communication, interpretation of clinical data, time management and the candidate's ability to handle typical interruptions whilst consulting. Pilot studies of cases designed to test a range of these skills have been carried out on behalf of the Royal College of General Practitioners and the results appear encouraging.^{70 71}

Assessing attitudes

One of the principal recommendations of *Tomorrow's doctors* is to inculcate attitudes of mind and behaviour that befit a doctor. If this objective is to be achieved, attitudes must also be assessed. However, in contrast to the situation described for clinical skills, there is a paucity of information regarding the assessment of attitudes. There are two possible alternatives, written assessment and assessment by observation.

It is possible to assess attitudes by written examination. The considerations regarding content specificity which were discussed earlier in relation to assessing knowledge will also apply to assessment of attitudes, thus short answer or modified essay-type questions would be recommended to ensure sufficient sampling. The example modified essay question presented earlier (box 3) would test both patient management and the attitudes of the candidate to patients.

The main object of including attitudes in the new curriculum is presumably to ensure that students develop the right attitudes in their future careers. One might argue that a written assessment of attitudes will not tell us how a student is going to behave in real-life. Observation would thus seem more attractive, and one could envisage objective structured clinical examination stations, objective structured long examination records, or simulated surgeries designed to test attitudes and behaviour. However, one must still ask the question 'Do students behave in the same way when they know that their attitudes are being assessed as they would in practice?', ie, does the act of being observed affect the act itself, a phenomenon referred to as the Hawthorne effect. Cynics would probably think not! It may be necessary to supplement such assessments with observation on rotations, ideally applying checklists.

Progress tests—weakening the link between assessment and learning

Discussion so far has assumed that assessment is used for summative purposes and recognises that assessment will drive the learning process. Progress testing is a novel assessment procedure designed to provide students and faculty staff with feedback about student progress, whilst not disrupting the learning process or driving learning styles.²⁵ Such tests cover a very wide spectrum of knowledge so it is not feasible to revise for them. As students progress through the course, their score on the progress test increases and they can compare their performance with the class means and standard deviations.

Evaluation of the progress test by questionnaire indicates that it provides useful feedback, that students do not study for the test and that it does not subvert the educational philosophies on which these curricula are based.^{14 72} Progress tests could form a useful element of assessment in revised curricula in the UK which incorporate problem-based learning.

Conclusion

Over the last few years there have been several exciting and interesting innovations in the assessment of medical students as a response to the deficiencies of the traditional methods of essays and oral examination. Most of these new methods are not used extensively in the UK and there is considerable scope for further development and evaluation in revised curricula. The extended matching items and key features questions formats have good psychometric properties and are promising for the assessment of higher cognitive knowledge-based objectives.

However, it is apparent that no single method is appropriate for assessing all aspects of medicine. A combination of techniques will thus be required to satisfy the GMC's requirements. One can envisage a multi-faceted assessment of knowledge, skills and attitudes, with objective structured clinical examination-type tests used throughout the course to certify competence in individual clinical skills, leading on to a well designed objective structured long examination record or stimulated surgery type examination in the later stages of the course, supplemented by observation and global ratings as appropriate. These would need to be complemented by objective examinations consisting of a mix of question formats to ensure testing of knowledge at higher cognitive levels.

There is thus a range of methods available for assessment, the challenge to medical schools is to utilise these to achieve the required changes in medical education. The next few years should produce interesting evaluations of assessment in the new curricula!

We are grateful to our colleagues on the Curriculum Management Team for useful discussions on assessment and for permission to use past examination questions.

- 1 Atkins MJ, Beattie J, Dockrell WB. *Assessment issues in higher education*. Sheffield: Department of Employment, 1992.
- 2 Wakeford R, Southgate L. Postgraduate medical education: modifying trainees' study approaches by changing the examination. *Teach Learn Med* 1992;4:210-3.
- 3 GMC. *Tomorrow's doctors: recommendations on undergraduate medical education*. London: GMC, 1993.
- 4 Harden RM, Davis MH. AMEE Medical Education Guide No 5. The core curriculum with options or special study modules. *Med Teach* 1995;17:125-48.
- 5 Parsell GJ, Bligh J. The changing context of undergraduate medical education. *Postgrad Med J* 1995;71:397-403.
- 6 Entwistle N, Entwistle A. Contrasting forms of understanding for degree examinations: the students experience and its implications. *Higher Educ* 1991;22:205-27.
- 7 Brown S, Knight P. *Assessing learners in higher education*. London: Kogan Page, 1994.
- 8 Newble DI, Jaeger K. The effect of assessments and examinations on the learning of medical students. *Med Educ* 1983;17:165-71.
- 9 Barrows HA, Williams RG, Moy RH. A comprehensive performance based assessment of fourth-year students' clinical skills. *J Med Educ* 1987;62:805-9.
- 10 van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardised patients: state of the art. *Teach Learn Med* 1990;2:58-76.
- 11 Vu NV, Barrows HS. Use of standardised patients in clinical assessments: recent developments and measurement findings. *Educ Res* 1994;23:23-30.
- 12 Verwijnen M, Imbos T, Snellen H. The evaluation system at the medical school of Maastricht. *Assess Eval Higher Educ* 1982;7:225-44.
- 13 Smith RM. The triple-jump examination as an assessment-tool in the problem-based medical curriculum at the University of Hawaii. *Acad Med* 1993;68:366-72.
- 14 Blake JM, Norman GR, Kinsey E, Smith M. Report card from McMaster: student evaluation at a problem-based medical school. *Lancet* 1995;345:899-901.
- 15 Westmoreland M, Parsons M. Triple jump exercise. In: Shannon S, Norman G, Falconer J, eds. *Evaluation methods: a resource handbook*. Hamilton, Canada: McMaster University, 1995; pp 33-6.
- 16 Harden RM. How to assess students: an overview. *Med Teach* 1979;1:65-70.
- 17 Lowry S. Assessment of students. *BMJ* 1993; 306:51-4.
- 18 Gronlund NE. Appendix E. Taxonomy of educational objectives (major categories and illustrative objectives). In: *Measurement and evaluation in teaching*, 5th edn. New York: Macmillan, 1985; pp 513-20.
- 19 Palmer D, Rideout E. Essays. In: Shannon S, Norman G, eds. *Evaluation methods: a resource handbook*. Hamilton, Canada: McMaster University, 1995; pp 59-66.
- 20 Day SC, Norcini JJ, Diserens D, et al. The validity of an essay test of clinical judgement. *Acad Med* 1990;65:S39-S40.
- 21 Premadasa IG. A reappraisal of the use of multiple choice questions. *Med Teach* 1993;15: 237-42.
- 22 Richardson R. The multiple choice true/false question: what does it measure and what could it measure? *Med Teach* 1992;14:201-4.
- 23 Swanson D, Case S. Trends in written assessment: a strangely biased perspective. In: Harden RM, Hart IR, Mullholland H, eds. *Approaches to the assessment of clinical competence*. Dundee: Centre for Medical Education, University of Dundee, 1992; pp 38-53.
- 24 Triska OH, Skakun EN, Maguire TO, Harley CH. Judges' standard setting behaviour and students' thought processes in multiple choice items. In: Rothman AI, Cohen R, eds. *The sixth Ottawa conference on medical education*. Toronto: University Bookstore, 1994; pp 227-31.
- 25 Norman G. Multiple choice questions. In: Shannon S, Norman G, Falconer J, eds. *Evaluation methods: a resource handbook*. Hamilton, Canada: McMaster University, 1995; pp 47-54.
- 26 Maguire T, Skakun EN, Triska OH. Student thought processes evoked by multiple choice and constructed response items. *The seventh Ottawa conference on medical education and assessment*. Maastricht, 1996; p 253.
- 27 Armstrong WG, Reed FB, Wells PJ. *Self-assessment in biochemistry for medicine and dentistry*. Oxford: Blackwell Scientific, 1984.
- 28 Joorabachi B. How to construct problem-solving MCQs. *Med Teach* 1981;3:9-13.
- 29 Case SM, Swanson DB. Extended matching items: A practical alternative to free-response questions. *Teach Learn Med* 1993;5:107-15.
- 30 Wakeford RE, Roberts S. Short answer questions in an undergraduate qualifying examination - a study of examiner variability. *Med Educ* 1984;18:168-73.
- 31 Knox JDE. How to use modified essay questions. *Med Teach* 1980;2:20-4.
- 32 McGuire CH. Assessment of problem-solving skills. *Med Teach* 1980;2:74-9.
- 33 Page G, Bordage G, Allen T. Developing key-features problems and examinations to assess clinical decision-making skills. *Acad Med* 1995;70:194-201.
- 34 Norman G, Bordage G, Curry L, et al. A review of recent innovations in assessment. In: Wakeford R, Bashook P, Jolly B, Rothman R, eds. *Directions in clinical assessment: proceedings of the first Cambridge conference*. Cambridge, 1985; pp 9-27.
- 35 Page G, Bordage G. The Medical Council of Canada's key features project: a more valid written examination of clinical decision-making skills. *Acad Med* 1995;70:104-10.
- 36 MacRury K. Testing key features of the clinical case: a written format for a pre-clerkship examination. *The sixth Ottawa conference on medical education*. Ottawa, 1992; pp 201-2.
- 37 Newble DI, Dauphinee D, Macdonald M, et al. Guidelines for assessing clinical competence. *Teach Learn Med* 1994;6:213-9.
- 38 van der Vleuten CPM, Newble DI. How can we test clinical reasoning? *Lancet* 1995;345:1032-4.
- 39 Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979;13:41-54.
- 40 Harden RM, Gleeson FA. ASME Medical Education Booklet No 8 Assessment of medical competence using an objective structured clinical examination (OSCE). *Med Educ* 1979; 13:39.
- 41 McFaul PB, Taylor DJ, Howie PW. The assessment of clinical competence in obstetrics and gynaecology in two medical schools. *Br J Obstet Gynaecol* 1993;100:842-6.
- 42 Matsell DG, Wolfish NM, Hsu E. Reliability and validity of the objective structured examination in paediatrics. *Med Educ* 1991;25:293-9.
- 43 Dissanayake AS, Ali BA, Nayar U. The influence of the introduction of objective structured practical examinations in physiology on student performance at King Faisal University Medical School. *Med Teach* 1990;12:297-304.
- 44 Woodburn J, Sutcliffe N. The reliability, validity and evaluation of the objective structured clinical examination in podiatry (chiroprody). *Assess Eval Higher Educ* 1996;21:131-46.
- 45 Roberts J, Norman G. Reliability and learning from the objective structured clinical examination. *Med Educ* 1990;24:219-23.
- 46 Hull AL, Hodder S, Berger B, et al. Validity of three clinical performance assessments of internal medicine clerks. *Acad Med* 1995;70: 517-22.
- 47 A-Latif A. An examination of the examinations: the reliability of the objective structured clinical examination and clinical examination. *Med Teach* 1994;14:179-183.
- 48 Selby C, Osman L, Davis M, Lee M. How to do it: set up and run an objective structure clinical exam. *BMJ* 1995;310:1187-90.
- 49 Reznick RK, Smee S, Baumber JS, et al. Guidelines for estimating the real cost of an objective structured clinical examination. *Acad Med* 1993;68:513-7.
- 50 Shatzer JH, Darosa D, Colliver JA, Barkmeier L. Station-length requirements for reliable performance-based examination scores. *Acad Med* 1993;68:224-9.
- 51 Colliver JA, Vu NV, Markwell SJ, Verhulst SJ. Reliability and efficiency of components of clinical competence assessed with five performance-based examinations using standardised patients. *Med Educ* 1991;25:303-10.
- 52 Hodder RV, Rivington RN, Calcutt LE, Hart IR. The effectiveness of immediate feedback during the OSCE. *Med Educ* 1989;23: 184-8.
- 53 Black NMI, Harden RM. Providing feedback to students on clinical skills by using the Objective Structured Clinical Examination. *Med Educ* 1986;20:48-52.
- 54 Blackwell TA, Ainsworth MA, Dorsey NK, Callaway MR, Rodgers LP, Collins KE. A comparison of short answers and extended matching question scores in an OSCE. *Acad Med* 1991;66:S40-S42.
- 55 Hodges B, Turnbull J, Cohen R, Bienstock A, Norman G. Evaluating communication skills in the objective structured clinical examination format: reliability and generalizability. *Med Educ* 1996;30:38-43.
- 56 Kinderman P, Humphris G. The role of cognitive schemata in clinical communication skills training. *Med Educ* 1995;29:436-42.
- 57 Gleeson F. The effect of immediate feedback on clinical skills using the OSCE. *The sixth Ottawa conference on medical education*. Ottawa: University of Toronto, Bookstore Custom Publishing, 1994; p 412-5.
- 58 Southgate L. Freedom and discipline: clinical practice and the assessment of clinical competence. *Assess Eval Higher Educ* 1994;44: 87-92.
- 59 Newble DI. Assessment of clinical competence: state of the art. In: Bender W, Hiemstra R, Scherpier A, Zwierstra R, eds. *Third international conference on teaching and assessing clinical competence*. Groningen, The Netherlands, 1990; pp 23-7.
- 60 Muzzin LR, Hart L. Oral examinations. In: Neufeld VR, Norman GR, eds. *Assessing clinical competence*. New York: Springer Publishing Company, 1985; pp 71-93.
- 61 Muzzin LJ. Oral examinations. In: Shannon S, Norman G, Falconer J, eds. *Evaluation methods: a resource handbook*. Hamilton, Canada: McMaster University, 1995; pp 37-42.
- 62 Evans L, Ingersoll RW, Smith EJ. The reliability, validity and taxonomic structure of the oral examination. *J Med Educ* 1966;41:651-7.
- 63 Porkorney AD, Frazier SH. An evaluation of oral examinations. *J Med Educ* 1966;41:28-40.
- 64 Waugh D, Moyse CA. Medical Education II: oral examinations: a video study of the reproducibility of grades in pathology. *Can Med Assoc J* 1969;100:635-40.
- 65 Wigton RS. The effect of student personal characteristics on the evaluation of clinical performance. *J Med Educ* 1980;55:423-7.
- 66 Wakeford R, Southgate L, Wass V. Improving oral examinations - selecting, training, and monitoring examiners for the MRCGP. *BMJ* 1995;311:931-5.
- 67 Streiner DL. Global rating scales. In: Neufeld VR, Norman GR, eds. *Assessing clinical competence*. New York: Springer Publishing Company, 1985; pp 119-41.
- 68 Streiner DL. Clinical ratings - ward evaluations. In: Shannon S, Norman G, Falconer J, eds. *Evaluation methods: a resource handbook*. Hamilton, Canada: McMaster University, 1995; pp 29-31.

- 69 Newble DI. The observed-long case in clinical assessment. *Med Educ* 1991;25:369–73.
- 70 Bingham L, Burrows P, Caird R, Holsgrove G, Jackson N, Southgate L. Simulated surgery: a framework for the assessment of clinical competence. *Educ Gen Pract* 1994;5:143–50.
- 71 Bingham L, Burrows P, Caird R, Holsgrove G, Jackson N. Simulated surgery - using standardised patients to assess the clinical competence of GP registrars - a potential clinical component of the MRCGP examination. *Educ Gen Pract* 1996;7:102–11.
- 72 Brummer I, Linden ML, Hoogenboom RJJ, Scherpbier AJJA, Verwinjen GM. The Maastricht progress test: the students' opinion. *The seventh Ottawa international conference on medical education and assessment*. Maastricht, 1996; p 52.

Medical Anniversary

(Lord) THOMAS HORDER, 7 JANUARY 1871

(Lord) Thomas Horder (1871–1955) was born at Shaftesbury, Dorset, where his father was a draper. He qualified at St Bartholomew's Hospital (1898) and became a consultant physician there (1912–1936) and at the Royal Northern Hospital (1899–1914). He was known as the man who brought the laboratory to the bedside because of his book entitled *Clinical pathology in practices*, published in 1910. In the same year he was called into consultation because of King Edward VII's glycosuria. He showed that the reduction of Fehling's solution, which had been attributed to glycosuria, was due to salicylates in a gargle which had been prescribed to the King (Medvei and Thornton 1974, The Royal Hospital of St Bartholomew). He was also physician to Edward, Prince of Wales, George VI, and Elizabeth II.

He died on 13 August 1955. A *British Medical Journal* obituary said "The death of Lord Horder removes the most outstanding clinician of his time... He was more than a physician at the head of his profession, he was an interpreter of medicine... a dispassionate thinker, never afraid of the truth, and able to employ language in such a way that... none could mistake his meaning... a good doctor, jealous for the honour of his profession and its freedom to serve the community".

— D G James