

Conclusions

Therapeutic and toxic plasma concentrations of commonly measured drugs

Drug	Concentration below which a therapeutic effect is unlikely	Concentration above which a toxic effect is more likely
Aminoglycosides:		
Amikacin	34 µmol/l (20 µg/ml) (at peak) 17 µmol/l (10 µg/ml) (at trough)	55 µmol/l (32 µg/ml) (at peak)
Gentamicin	5 µg/ml (at peak) 2 µg/ml (at trough)	12 µg/ml (at peak)
Kanamycin	50 µmol/l (25 µg/ml) (at peak) 20 µmol/l (10 µg/ml) (at trough)	80 µmol/l (40 µg/ml) (at peak)
Cardiac glycosides:		
Digitoxin	20 nmol/l (15 ng/ml)	39 nmol/l (30 ng/ml)
Digoxin	1.0 nmol/l (0.8 ng/ml)	3.8 nmol/l (3 ng/ml)
Cyclosporin*	80-200 nmol/l (100-250 ng/ml)	170-330 nmol/l (200-400 ng/ml)
Lithium	0.4 mmol/l	1.0 mmol/l
Phenytoin	40 µmol/l (10 µg/ml)	80 µmol/l (20 µg/ml)
Theophylline	55 µmol/l (10 µg/ml)	110 µmol/l (20 µg/ml)

*Measured in whole blood by specific radioimmunoassay or high performance liquid chromatography. The actual results depend on the laboratory in which the measurement is made.

In this series we have outlined the uses of measuring the plasma concentrations of some drugs and given guidelines on how such measurements should be made and interpreted.

The box summarises the target plasma concentrations for each of the drugs. In each case there is a concentration below which a therapeutic effect is unlikely and a concentration above which the risk of toxicity is high. These two concentrations imply a therapeutic range for each drug, but remember that there are circumstances in which strict adherence to a range of this kind is inappropriate. The plasma concentration should always be interpreted in the light of factors which may alter the effective therapeutic range.

Nor is it always necessary to measure plasma concentrations to achieve satisfactory drug therapy. Routine measurement without a clear purpose is as bad as no measurement at all. The application of the principles we have outlined should allow the rational use of plasma concentration measurement in optimising drug therapy.

- 1 Vozeh S. Cost-effectiveness of therapeutic drug monitoring. *Clin Pharmacokin* 1987;13:131-40.
- 2 Spector R, Park GD, Johnson GF, Vesell ES. Therapeutic drug monitoring. *Clin Pharmacol Ther* 1988;43:345-53.
- 3 McInnes GT. The value of therapeutic drug monitoring to the practising physician—an hypothesis in need of testing. *Br J Clin Pharmacol* 1989;27:281-4.
- 4 Sjöqvist F. Interindividual differences in drug responses: an overview. In: Rowland M, Sheiner LB, Steimer JL, eds. *Variability in drug therapy*. New York: Raven Press, 1985:1-10.
- 5 Larkin JG, Herrick AL, McGuire GM, Percy-Robb IW, Brodie MJ. Antiepileptic drug monitoring at the epilepsy clinic: a prospective evaluation. *Epilepsia* 1991;32:89-95.

Dr J K Aronson is clinical reader in clinical pharmacology, and Dr D J M Reynolds is clinical lecturer in clinical pharmacology, Radcliffe Infirmary, Oxford.

Medical Education

Assessment of students

Stella Lowry

This is the fifth in a series of articles examining the problems in medical education and their possible solutions

Student assessment is often described as “the tail that wags the dog” of medical education. It is seen as the single strongest determinant of what students actually learn (as opposed to what they are taught) and is considered to be uniquely powerful as a tool for manipulating the whole education process. Sir William Osler summed up the power of examinations in 1913: “At the best means to an end, at the worst the end itself, they may be the best part of an education or the worst—they may be its very essence or its ruin.”¹ But is assessment as powerful as we think, and, if it is, are most medical educators using it effectively?

Why assess?

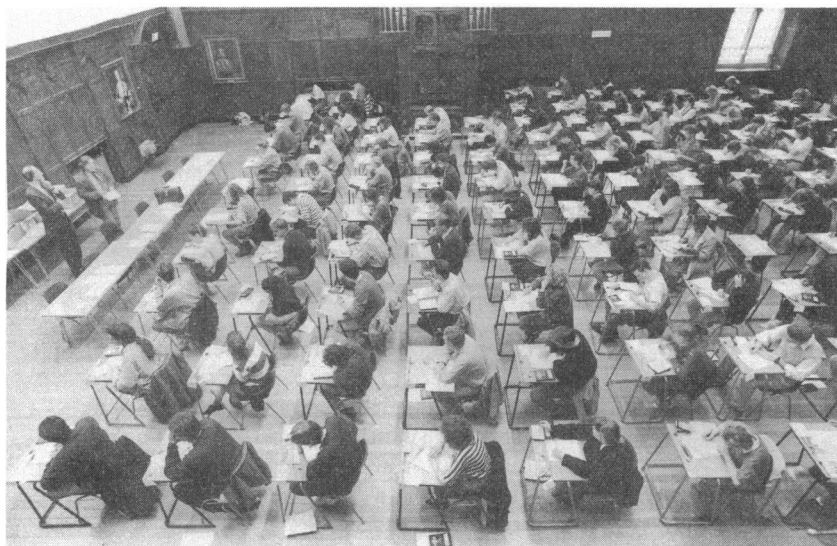
Few people formally question why we assess medical students, and many who do think no further than using assessment as a means of checking that required information has been learnt. Certainly in an overloaded curriculum students will pay attention to topics that they know will feature in examinations.² A recent study

of surgical students at the Flinders University of South Australia found that when no clear guidelines and course objectives were given in a self directed learning programme the students—far from exploring the topic widely and pursuing personal interests—tried to “guess” what would feature in the final examination and concentrated on that (D J Prideaux, paper presented at fifth Ottawa international conference on assessment of clinical competence, Dundee, September 1992). This tendency allows staff to direct students’ attention to important topics but also increases the risk that unexamined areas will be ignored.

Unfortunately, the fact that students can successfully answer examination questions on a topic is no guarantee that they will retain their knowledge of the subject. Assessments that are based on a one off factual recall are notoriously unreliable as indicators of real learning,³ and if assessment is to be used to ensure learning more complex approaches are needed. One method is to retest the same information at regular

British Medical Journal,
London WC1H 9JR
Stella Lowry, assistant editor

BMJ 1993;306:51-4



Tail wagging the dog of medical education. Medical students sitting finals at end of their course

UNIVERSITY OF NEWCASTLE UPON TYNE

separate formative tests are also available. It is inexcusable that many students who have failed an examination are still expected to retake the whole topic with no idea of where they went wrong last time and no guidance on where to focus their attention. That is a waste of their time and is not educational.

Ensuring minimum standards

Because undergraduate medical education is a professional training the minimum standards of professional practice must be acquired during the course. The protection of the public demands this gatekeeping function of assessment, even if it places constraints on the educational experience.

This requirement is universal. McMaster in Canada has no summative assessment during the course,⁶ but any graduate who wishes to practise medicine must still pass the national licensing exams. Indeed, one criticism often made of the McMaster course is that its students do not do as well as those from other schools in the licensing exams. Dr Barbara Ferrier, a founder member of the medical faculty at McMaster, thinks that this is an oversimplification. For one thing the relative ranking of McMaster students varies from year to year, and over several years no consistent pattern emerges. McMaster students tend to do a bit worse

intervals. At the University of Maastricht, for example, the entire medical school sits an identical multiple choice "progress test" at various points during the course.⁴ This allows students to document their increasing knowledge as they progress through the course and ensures that essential information is not conveniently forgotten once the exam is out of the way.

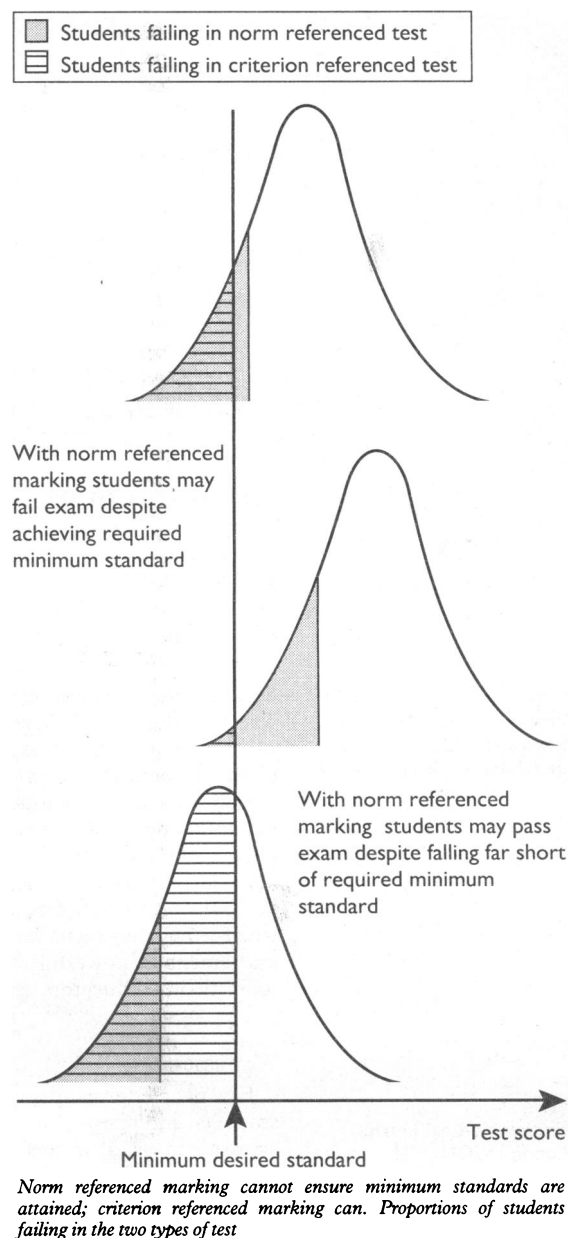
McMaster University in Canada, which has never had formal examinations during the medical course, is now planning to introduce a Maastricht-style progress test. This is largely in response to frequent complaints from students that the complete lack of formal assessment actually adds to their stresses rather than reduces them. Students never quite know whether they have learnt enough about a subject, and it is hoped that a progress test will give them some idea of when it is safe to stop.

Assessment as an aid to learning

An important function of assessment is to aid learning by providing students with a check on their progress and an opportunity to improve. If students are given an opportunity to give feedback to their tutors assessment can also be useful in refining the way a curriculum is taught and ensuring that course objectives are met. Too often assessment is used purely as a tool for staff to regulate progression of students through the system and rank them in order of achievement.

We tend to rely heavily on summative assessment—testing acquired knowledge at the end of a course (when it is too late to correct deficiencies). Students need formative assessment—regular checks on how they are doing with detailed feedback on the results and an opportunity to try again after remedial learning. This is a powerful tool in focusing students' learning⁷ and can be tailored to individual needs and wants. The Centre for Medical Education, University of Dundee, for example, has devised computer banks of multiple choice questions that students can use as and when they wish to monitor their own progress. Simple feedback can be incorporated into such programmes. More complex feedback may have implications for staff time, but if our aim is to teach rather than to test that must be accepted.

Some medical schools still refuse to provide any feedback on their examinations on the grounds that they need to use the test again the next year and mustn't allow the answers to leak out. A bank of three papers used in rotation should get round this problem, and if that is impossible then by all means hold a summative test without feedback but ensure that



than average on the multiple choice question paper and rather better than average in the patient management section. Secondly, McMaster allows all of its students to enter the licensing exams as soon as they complete the course whereas many other schools screen out those they think might fail and allow them to enter only after further tuition.

The only sensible way to ensure that minimum standards are met is to define them and fail any student who does not reach them. Why, then, do so many medical schools persist in marking students against their peers rather than against an externally set standard? Most British medical schools use so called norm referenced assessment, in which the students' scores in an exam are ranked on a normal distribution and the pass mark is adjusted to provide a predetermined percentage failure rate. This means that the absolute standard achieved may vary widely from year to year (figure). Because most medical students manage to learn the absolute essentials of a course these basic points have little discriminatory power in a norm referenced examination and are often omitted. The exams thus concentrate on the rarities that will discriminate between the best and worst students, and the basic principles may never be tested.

Norm referenced assessment also raises the possibility that a particularly poor group of students will "pass" the exam despite falling far short of the desired minimum achievement. Conversely, an excellent adaption in the way a course is taught may have dramatic effects on the amount that students learn about a subject, but this will not be reflected in any improvement in the "pass rate."

In criterion referenced assessment, however, the required minimum attainment is determined in advance and students who score less than this standard fail. Norm referenced marking cannot ensure that minimum standards are attained; criterion referenced assessment can. Norm referenced assessment is useful when scores are being ranked before allocation of a

Box 1

Objective structured clinical examination

Written papers may be useful tests of factual knowledge but are poor at assessing clinical skills. Traditional clinical examinations are difficult to standardise, and how well students perform may depend largely on the "luck of the draw" in determining who examines them, on which patients, and using what questions. The objective structured clinical examination was devised to provide a more standardised way of assessing clinical competence.^{15,16}

- Objectives of the test are identified and recorded. Objective structured clinical examination is then designed to cover all of required aspects
- Candidates rotate around series of "stations," at each of which they are asked to perform clinical task or answer questions on material provided
- Assessors are present at relevant stations to assess candidate's performance using standardised checklist
- Clinical models and simulated patients can be used during objective structured clinical examinations to allow large numbers of students to be tested on same clinical problem without causing fatigue or distress to real patients
- Marking can be completed as the objective structured clinical examination proceeds and prompt feedback is possible (some centres provide immediate feedback between each station)
- Objective structured clinical examinations can be expensive and administratively cumbersome to set up but are easy to mark and allow testing of skills that more traditional methods ignore.

Box 2

McMaster "triple jump" test

An important aspect of education is learning how to learn. Many medical courses now use large elements of "self directed learning" so that students become familiar with the skills and resources needed to keep up to date with the rapid changes in medical knowledge. The "triple jump" test was devised as a means of assessing students' competence at self directed learning.

- *Step 1:* Student reads written "problem" and discusses first impressions with tutor. Student then selects some tasks for further learning and decides what additional information is needed
- *Step 2:* During period of private study the student uses any relevant sources of information to tackle these self selected tasks
- *Step 3:* Student reports back to tutor and presents revised summary of problem based on his or her recent researches. Tutor provides feedback on way in which student has tackled problem.

limited commodity, such as house jobs, but should not be the main system for assessing medical students' achievements.

What should we assess?

If assessment is so important what should be assessed? Traditionally the emphasis in medical education has been on acquiring a body of essential facts on each topic, but now more attention is given to the skills and attitudes thought to be important in a "good doctor."^{7,9} No single examination can be expected to assess such a wide range of features. Medical examiners should identify those aspects that they wish to test and then provide a range of appropriate formats.

There is often a high correlation between students' performance in different types of examination, which has led some people to conclude that they can rely on whatever system is cheap and administratively simple. Professor David Newble, of the University of Adelaide, however, emphasises that this correlation may simply be reflecting the relative ability of students to study for exams rather than saying much about their actual knowledge. Most experts now agree that a range of methods is needed to assess a range of clinical skills.¹⁰ The multiple choice question paper, for example, is an excellent, cheap, reliable, and reproducible way of testing factual knowledge but tells us almost nothing about clinical skills. The objective structured clinical examination (box 1) is expensive, administratively cumbersome, and lacks high reliability but is an excellent way of assessing practical skills.¹¹⁻¹⁴ Even complex processes like skill at self directed learning can be tested; McMaster has devised the "triple jump test" to do just that¹⁷ (box 2).

Features of a good test

Given the plethora of techniques for assessing students, there is a real danger of becoming obsessed with the methods and neglecting the content. Professor Cees van der Vleuten, from Maastricht, reminded delegates at a recent conference on the assessment of clinical competence that ultimately the care with which any examination is devised is more important than the form that is chosen (paper presented at fifth Ottawa international conference on assessment of clinical competence, Dundee, September 1992). What, then, are the cardinal features of a good test?

A good test must be acceptable to those using it, feasible, valid, and reliable. A test may be acceptable to

"Excellent way of assessing practical skills." Objective structured clinical examination in progress at St Bartholomew's Hospital, London



ST BARTHOLOMEW'S HOSPITAL

some of those dealing with it and not to others. Multiple choice tests, for example, may be acceptable to those administering them because they are simple to use, cheap to run, and quick and easy to mark. They may be wholly unacceptable to those sitting them if the questions do not seem to be a fair test of the important aspects of a course. A carefully thought out objective structured clinical examination may be a good way to assess essential clinical skills, but the logistic problems of running it may make it unfeasible for the simultaneous assessment of an entire year of students.

A valid test measures what we want it to measure and nothing else. The validity of a test can be assessed under various headings (see glossary).¹⁸

Reliability is a measure of the consistency and accuracy with which a test measures what it is supposed to. In a good assessment system not only should the test be reliable but so also should the system for marking it. A multiple choice question paper is usually very reliable and feasible, but because it essentially tests factual recall it may not be very acceptable to candidates and would be unlikely to have high validity as a measure of clinical skills.¹⁹ Essay questions tend to have low reliability (particularly because of difficulties in standardising the marking) and surprisingly low validity. Structured short answer papers have higher validity because more precise instructions can be given to the candidate.²⁰

No single examination format will guarantee acceptability, feasibility, validity, and reliability, but care in identifying the strengths and weaknesses of each approach and clear objectives for the assessment should help staff select a useful range of tests.

Conclusions

Assessing medical students should help them focus their learning during the course, identify individual strengths and weaknesses, provide an opportunity for improvement, highlight deficiencies in the content or delivery of the medical course, and, ultimately, protect the public against incompetent graduates. To do all this the assessment system must contain a large and properly managed formative element. The summative assessment must be criterion referenced. There must be adequate feedback between staff and students to ensure that all potential gains are obtained from the assessment system.

Glossary

Summative assessment: Testing acquired knowledge at the end of a course.

Formative assessment: Regular testing of progress throughout a course.

Norm referenced tests: Scores are ranked on a normal distribution and the pass mark adjusted to achieve a predetermined percentage failure rate.

Criterion referenced tests: Pass mark is predetermined, and the pass rate is allowed to vary with absolute achievement.

Validity: Extent to which a test measures what we want it to measure and nothing else.

Content validity: Measure of whether the test contains a representative sample of the items that we wish to assess.

Construct validity: How well a test measures the feature of interest. (If students do better on a test after studying a course aimed at teaching a particular skill we have some evidence of the construct validity of that test for assessing that skill.)

Criterion referenced validity of a test is judged by assessing how well the test results correlate with another accepted assessment method.

Reliability: Measure of the consistency and accuracy with which a test measures what it is supposed to.

- Osler W. Examinations, examiners, and examinees. *Lancet* 1913;iii:1047-59.
- Newble DI, Jaeger K. The effect of assessments and examinations on the learning of medical students. *Med Educ* 1983;17:165-71.
- Tooth D, Tonge K, McManus IC. Anxiety and study methods in preclinical students: causal relation to examination performance. *Med Educ* 1989;23:416-21.
- Verwijnen GM, Imbos T, Snellen H, Stalenhoef B, Pollenmans M, Luyck S, et al. The evaluation system of the medical school of Maastricht. *Assessment and Evaluation in Higher Education* 1982;3:225-44.
- Ende J. Feedback in clinical medical education. *JAMA* 1983;250:777-81.
- Pallie W, Carr DH. The McMaster medical education philosophy in theory, practice and historical perspective. *Med Teach* 1987;9:59-71.
- Turnbull JM. What is... Normative versus criterion-referenced assessment? *Med Teach* 1989;11:145-50.
- General Medical Council Education Committee. *Recommendations on basic medical education*. London: GMC, 1980.
- Towle A. *Critical thinking: the future of undergraduate medical education*. London: King's Fund Centre, 1991.
- Jolly B, Wakeford R, Newble D. Implications for action and research. In: Newble D, Jolly B, Wakeford R, eds. *Certification and recertification in medicine*. Cambridge: Cambridge University Press (in press).
- Petrusa ER, Blackwell TA, Rogers LP, Saydjari C, Parcel S, Guckian JC. An objective measure of clinical performance. *Am J Med* 1987;83:34-42.
- Newble DI. Eight years' experience with a structured clinical examination. *Med Educ* 1988;22:200-4.
- Malik SL, Manchanda SK, Deepak KK, Sunderam KR. The attitudes of medical students to the objective structured practical examination. *Med Educ* 1988;22:40-6.
- Roberts J, Norman G. Reliability and learning from the objective structured clinical examination. *Med Educ* 1990;24:219-23.
- Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *BMJ* 1975;ii:447-51.
- Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination. *Med Educ* 1979;13:41-54.
- Painvin C, Neufeld V, Norman G, Walker I, Whelan G. The triple jump exercise—a structured measure of problem solving and self-directed learning. *Annual Conference on Research in Medical Education* 1979;18:73-83.
- Newble DI, Hoare J, Elmslie RG. The validity and reliability of a new examination of the clinical competence of medical students. *Med Educ* 1981;15:46-52.
- Levine HG, McGuire CH, Nattress LW. The validity of multiple choice achievement tests as measures of competence in medicine. *American Educational Research Journal* 1970;7:69-82.
- Felett GI, Engel CE. The modified essay question for testing problem-solving skills. *Med J Aust* 1980;ii:79-80.