

# Formative (in-course) graduate assessment

## Design of an accurate multiple-choice question examination and the training merits thereof

R. F. GLEDHILL

### Summary

Medical graduate learning is being appraised to an increasing extent during training. The adequacy of these formative (in-course) assessments has not been widely studied.

The performances and opinions of junior specialists were used to evaluate the accuracy of a graduate multiple-choice question examination (MCQE) and the merits of such in-course assessments. Methods used in verifying the validity and reliability of this MCQE are detailed. In-course assessments were rated as being valuable in graduate training.

Issues pertinent to sound appraisals of graduate learning are discussed.

*S Afr Med J* 1983; 64: 476-477.

In keeping with the view that specialist competence should be examined at all stages of training,<sup>1-3</sup> the Department of Internal Medicine at Stellenbosch University has recently introduced formative (in-course) assessments at the end of each 6-month rotation in the 4-year programme, which includes a rotation in neurology. The neurology staff considered that there were persuasive arguments<sup>1,4-6</sup> for including a multiple-choice question examination (MCQE), not only to ensure a fair test covering a broad range of facts and concepts but also to uncover defects in their own methods of teaching.<sup>5</sup>

The adequacy of in-course assessments has not been the subject of widespread study. If specialty competence is to be judged with precision, it is important that the relevant tests should accurately sample and measure professional abilities<sup>3</sup> — examinations can provide adequate information regarding competence only when the methods used are reliable and valid, which they often are not.<sup>1</sup> Furthermore, graduates should perceive that the appraisals are beneficial to their training. The purpose of this report is twofold: (i) to illustrate how the validity and reliability of an in-course graduate MCQE can be verified; and (ii) to record graduate evaluations of in-course assessments.

### Methods

The MCQE (in applied basic neurosciences and clinical neurology) comprised 62 questions of the grouped any-from-five

type,<sup>7</sup> the complexity and correctness of answers having been determined by reference to authoritative publications.<sup>6</sup> Instructions advised that negative marking<sup>8</sup> for wrong answers would be used to neutralize the effect of guessing. Space was provided for comments on the intelligibility of each question and for recording completion time, for which no limit was imposed.

An accompanying six-part questionnaire consisted of five items concerning the validity of the examination, its value as an educational and learning exercise, and the acceptability of the MCQE format (see Table I); possible answers of 'yes', 'neutral' and 'no' were rated scores of 2, 1 and 0 respectively. Part six called for comments on the composition of the examination. All 7 physicians who had been awarded the M. Med. degree and/or F.C.P. within the previous 9 months (4 in 1981, 1 with distinction, and 3 in 1982) and the single neurology registrar (final year) agreed to participate.

In order to judge objectively the merits of any test it is necessary to determine its validity and reliability.<sup>9</sup> Since there exists no simple or single measure, evidence of content, construct and criterion-related test validity must be provided.<sup>9</sup> *Content validity* is a measure of the degree to which the test contains a representative sample of desired competences;<sup>9</sup> the opinions of recently specialized physicians should provide a suitable and relevant estimate here. *Construct validity* determines how well performances measure the attributes upon which the examination is based<sup>9</sup> (those physicians who obtained a pass grade would be expected to perform less well than the 1 who gained a distinction, and all 7 less well than the neurology registrar). *Criterion-related concurrent validity* relates test results to an independent, valid and reliable assessment of current performance<sup>9</sup> — the recent success in M. Med. and F.C.P. examinations gained by our subjects should fulfil these requirements.

No test can be valid unless it is also reliable.<sup>9</sup> In theory, a reliable test should produce the same result if administered to the same person on two separate occasions, the degree to which it does so being indicated by the coefficient of stability.<sup>9</sup> Practical considerations demand that estimates of stability be provided by measures of internal consistency, in which individual scores in the first half of the test are correlated with those of the second half. The coefficient of reliability (R) is calculated from the equation  $R = \frac{2r}{1+r}$  (Spearman-Brown), where  $r$  = the correlation coefficient measuring internal consistency. Estimates of probability were calculated using the standard  $t$  test, with  $P < 0,05$  indicating significance.

### Results

The examination was completed in a mean time of 82,5 minutes (range 45 - 125 minutes) with an average of 1,3 minutes per question. There was a significant correlation between time taken and score achieved ( $r = 0,789, P = 0,02$ ). The individual scores of the 8 graduates were as follows: the 6 'pass grade' physicians scored 42%, 43%, 45% (2), 47% and 54%, the physician with distinction scored 64% and the neurology registrar 80%. Aggregated ratings of the 7 physicians' answers to the

---

Departments of Internal Medicine and Neurology, University of Stellenbosch and Tygerberg Hospital, Parowvallei, CP  
R. F. GLEDHILL, B.S.C., M.B. B.S., M.R.C.P., M.D. Senior Lecturer and Principal Specialist (Present address: Little Roke, Bouverie Road, Chipstead, Coulsdon, Surrey, England)

---

TABLE I. MCQE QUESTIONNAIRE WITH AGGREGATE SCORES OF THE 7 PHYSICIANS' RATINGS

Item	Rating (maximum=14)	Score (%)*
Do you consider that the questions were valid, in terms of:		
Familiar fundamental concepts in neurology?	11	78,5
Competences necessary to pass M.Med. or F.C.P.?	11	78,5
Practical requirements for an effective physician?	11	78,5
Do you consider that such an examination has educational value?	12	85,5
If this examination had been held during your neurology rotation, would you have undertaken more self-education in neurology?	10	71,5
Do you prefer the MCQE format to an essay-type examination?	9	64,5
Do you approve of the holding of a written and practical examination assessment on completion of the neurology rotation?	13	93,0

\*To nearest 0.5%.

questionnaire on the MCQE are listed in Table I. Of the 3 items estimating content validity, there was greatest agreement (5/7) that 'competences' necessary to pass specialist diplomas were tested. Four graduates expressed preference for the MCQE format and 2 for essay-type examinations.

Comments on the composition of the examination concentrated on two main themes: the questions were too detailed and complex for physicians and the topics too restricted in scope. Criticisms of question intelligibility totalled 18, 16 of which were justifiable. The coefficient of reliability of the examination was highly significant ( $R = 0,9685$ ,  $P < 0,0001$ ).

## Discussion

The first prerequisite of a valid examination is that the content must be relevant and appropriate to the competence level required.<sup>9</sup> While the judgements of physicians who had recently specialized were considered suitable for such an appraisal, it would clearly have been preferable to sample more individual opinions. It might also be argued that after only a few months a physician is not adequately experienced to make this type of judgement. However, these were the only opinions readily available since Stellenbosch graduates in internal medicine tend to leave soon after specializing; indeed, within 1 month of completing this study 2 of the 7 physicians had left. Other investigators may be more fortunate in this respect. However, in order to ensure construct and criterion validity this interval should not be unlimited since there is evidence<sup>3</sup> of a decline in competence with time. Despite being rated highly for overall validity, the examination content was generally thought to be too restricted, a shortcoming likely to be prevalent when subspecialists are responsible for the design. This type of information is invaluable if the introduction of assessments is to fulfil more than sectional interests.

Criticisms that the questions were too complex were reflected in the scores, which nevertheless demonstrated the ability of the test to discriminate between the more and the less competent, thereby providing evidence of satisfactory construct validity. While calculating an 'easiness index'<sup>8</sup> allows exclusion of the most difficult questions, this is likely to affect content validity. The most satisfactory solution will be to devise easier questions on the same topics. Despite the high degree of test reliability, the number of questions that were considered either ambiguous or imprecise emphasizes the need to test and discuss<sup>10</sup> an MCQE before the formal use thereof.

Whereas the average of 1,3 minutes for completion of each question is in accord with MCQE recommendations,<sup>7</sup> the significant relationship between time taken and score achieved suggests that time limits may reduce test accuracy.

The issue of the relative merits of essay- and objective-type tests<sup>4</sup> (often a matter of contention not based on solid data<sup>3</sup>) is beyond the scope of this report, but a properly designed MCQE successfully tests that knowledge without which a doctor cannot

be really effective.<sup>6</sup> The preferences recorded by physicians in the present study must be considered in the context of their previous tertiary education and the present tendency for final internal medicine graduate examinations in South Africa to be dominated by essay-type questions. One physician suggested that formative and summative tests should be similar in type. Much as obsession with examination success is decried because of its negative educational effect, this attitude will remain for as long as a single, summative 'pass or fail' examination can determine a graduate's future career.

Although 2 physicians considered that in-course assessment would not have influenced their self-education in neurology, the overall highly favourable response to its introduction, educational value and motivating effect is an encouragement to those who might question the desirability of the additional obligation.

## Conclusion

The design of a valid and reliable in-course graduate MCQE is time-consuming and critically dependent on the goodwill and integrity of colleagues. It is likely that even greater demands on time and resources will be made in designing a valid and reliable practical examination. However, if in-course specialist assessment is to provide a proper appraisal of progress and, in particular, to judge with precision suitability for continuing training, the relevant tests must be shown to estimate competence accurately.

Graduates consider that in-course assessments could benefit their learning; it is incumbent upon teachers to reciprocate by designing tests accordingly.

I am indebted to Dennis Capatos for reviewing the statistical analyses and for criticisms of the manuscript.

## REFERENCES

- Anonymous. Assessment in medical education. *Med Educ* 1978; 12: 265-266.
- Royal Commission on Medical Education, 1965-1968. London: HMSO, 1968: 51-53.
- Miller GE. The orthopaedic training study. *JAMA* 1968; 205: 601-606.
- Cox KR. What type of written examination should I use? In: Cox KR, Ewan CE, eds. *The Medical Teacher*. Edinburgh: Churchill Livingstone, 1982: 197-199.
- Lennox B, Anderson JR, Moorhouse P. A comparative trial of objective papers and essay papers in pathology and bacteriology class examinations. *Lancet* 1957; ii: 396-402.
- Smart GA. The multiple choice examination paper. *Br J Hosp Med* 1976; 15: 131-136.
- Harden RM. Constructing multiple choice questions of the multiple true/false type. *Med Educ* 1979; 13: 305.
- Lennox B. *Hints on the Setting and Evaluation of Multiple Choice Questions of the One-from-five Type* (booklet No. 3). Dundee: Association for the Study of Medical Education, 1974.
- 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.
- Cox KR. How to construct a fair multiple choice question paper. In: Cox KR, Ewan CE, eds. *The Medical Teacher*. Edinburgh: Churchill Livingstone, 1982: 211-214.