Summative assessment of vocational trainees: results of a 3-year study

L M CAMPBELL

T S MURRAY

SUMMARY

Background. *Summative assessment of general practice trainees will be introduced across the UK for trainees completing vocational training after 1 September 1996.*

Aims. *A study was set up to assess the reliability of the consulting skills assessment which will be used to develop audit and factual knowledge testing as summative assessment tools and to determine the probable impact of the system on the numbers of trainees identified as not yet competent for independent practice.*

Method. *Videotaped consultations, an audit project, a trainer's report and a multiple-choice paper were evaluated for 359 trainees in the West of Scotland completing their training between 31 July 1993 and 31 July 1995.*

Results. *A total of 77 trainees (22%) were identified as being potentially of doubtful competence and 17 (5%) were adjudged to be as yet incompetent for independent practice. Videotaped consultation produced the highest pickup rate (14) followed by trainer's judgement (3). Subsequent analysis of the video tapes identified one trainee who should have been picked up by this method but was not.*

Conclusion. *The videotape instrument reliably identified the non-competent trainees. Audit and factual knowledge tests have a distinct role in summative assessment. The proportion of trainees being refused certificates by this assessment is likely to increase considerably but will still represent only about 5% of all trainees.*

Keywords: summative assessment; vocational trainees; consulting skills.

Introduction

THE Joint Committee for Postgraduate Training in General Practice has stated that a system of summative assessment of vocational trainees will apply to all general practitioner trainees completing their training after September 1996.[1] The reasons given for this decision were that the public have a right to expect all general practitioners to be competent to practice, and that there is evidence (although only anecdotal) that the competence of some newly qualified general practitioners is doubtful. The components chosen to form the assessment process in the West of Scotland region have already been described:[2] a multiple-choice paper, an assessment of videotaped consultations, an audit project and the trainer's judgement. The video assessment instrument has been selected by the UK regional advisers for use on a national basis.[3]

Trainees who do not reach a satisfactory standard in any of the four components are initially the subject of a review involving the regional adviser, the associate adviser (the equivalent of the course organizer in England) and the trainer. Thereafter, if concern persists regarding a trainee's competence, he or she is referred for further review to two assessors, who make recommendations as to whether a further period of training is required. When the system becomes nationwide, this extra pair of assessors will be taken from a national panel of trained assessors. In the UK system, the multiple-choice (true/false) paper will not lead to a referral process but will produce a pass or fail result.

An initial pilot study assessed the reliability and feasibility of videotaped consultations and described an instrument for assisting assessors in judging trainee competence.[4] This paper describes a full-scale study in which all trainees completing their trainee year between 31 July 1993 and 31 July 1995 participated. The decision concerning whether to issue a certificate of satisfactory completion remained the prerogative of the trainers but the decision was expected to be guided by the results of summative assessment.

The aims of this study were to assess further the reliability of the video assessment process, to determine the role of the audit project and multiple-choice paper, and to obtain an idea of the impact of the system on the numbers of trainees requiring further training.

Method

The general practice component of training was completed by 359 trainees during the study period. All of the trainees participated in the study, 343 taking part in all components.

Trainer's judgement of competence

All trainers were asked to complete a statement giving their judgement on the competence of the trainees under their tutelage. Trainers of the first group of trainees assessed (those who finished the course in July 1993) were offered a choice of three statements to make about the trainees:

(1) The trainee is competent to carry out the work of general practice
(2) The trainee is of doubtful competence
(3) The trainee is not competent and will require further training.

If a response of (2) or (3) was given, the trainer was asked to give details of their concerns.

For later batches of trainees, the trainer's statement was modified as shown in Table 1. This revised version was developed at a workshop run by one of the authors (LMC) involving trainers from Nottinghamshire and was circulated to all trainers in the West of Scotland region with a request for comment. All trainers found the form acceptable and suggested no major changes.

Multiple-choice paper

Each trainee sat a multiple-choice paper, without negative marking. A pass mark was derived using the Angof technique[5] as follows. A group of experienced general practitioner principals analysed the paper question by question, and for each question, produced a figure for the percentage of trainees whom they would expect to answer the question correctly. By this means, a

L M Campbell, FRCGP, assistant adviser in assessment; and T S Murray, FRCGP, regional adviser in general practice, Department of Postgraduate Medical Education, University of Glasgow.
Submitted: 7 October 1995; accepted: 17 November 1995.

Table 1. Revised form for trainer's judgement component of assessment of competence of trainees.

Clinical competence: Do you have any doubts about your trainess competence to perfrom unsupervised general practice? Yes/No
Professionalism: Does that trainee behave responsibly with regard to the areas of confidentiality, continuing education, and relationships with colleagues, staff and patients? Yes/No/Doubtful
Reliability: Does the trainee behave in a responsible manner with regard to duties within the practice such as surgeries and home visits? Yes/No/Doubtful
Personal organization: Do you think the trainee will be able to cope with the organizational difficulties of general practice, particularly time management? Yes/No/Doubtful
Other areas: Are there any other areas in which you have doubts about the trainee's fitness for independent practice? Yes/No/Doubtful
If yes, please specify.

minimum acceptable overall score was determined. Any candidate scoring more than one standard error below this score was entered into the referral process. The purpose behind this apparently complex system was to avoid rank ordering, which would result in failure of a predetermined percentage, and thus, run counter to one of the basic principles of summative assessment: that it should be possible for every trainee to pass the assessment process.

Audit

All audits were reviewed by two individuals with audit experience. The criteria used were:

- The aim of the audit should be clear
- The criteria and standards should be clear
- The measurements should be repeatable
- The cycle should be completed if possible
- The audit should be of educational value
- Appropriate proposals for change should be put forward.

If an audit satisfied these criteria, it was deemed to be acceptable; if not, the trainee was asked to modify and re-submit the audit. The criteria used are those widely agreed in the literature on audit.[6] Evidence of change brought about by the audit was not required, since instituting change was outwith the trainee's control.

Videotape assessment

The criteria used[2] and the reliability of the instrument[4] have been discussed elsewhere, but in summary, each trainee was assessed by two assessors working independently. If either assessor rated the trainee as 'unsatisfactory' or 'of doubtful competence', the trainee was entered into the referral process.

Our previous work suggested[4] that this process would identify 95% of those trainees with inadequate consulting competence, while also identifying a total of 20% where further assessment would indicate that the majority were competent, i.e. the anticipated sensitivity of the process was 95% with 80% specificty.

Results

Trainer's judgement

Trainers' reports were received for 359 trainees, three of whom were rated as unacceptable.

Multiple-choice papers

Out of the 359 trainees, 358 completed these papers between March 1993 and May 1995. For geographical reasons, the missing trainee had attended a different region for the day-release course and consequently missed the opportunity to sit the paper. As the summative assessment process developed, multiple-choice papers were chosen from three different sources:

- those sitting the paper in May 1993, December 1993 and December 1994 (145 trainees) took a multiple-choice paper developed within the region
- all trainees in the West of Scotland region in the last 6 months of their course sitting in May 1994 took the paper set by the Royal College of General Practitioners, as part of a joint initiative, including those not taking the MRCGP examination
- trainees sitting the examination in March 1995 took a paper based on the question bank of the fellowship examination of the Royal Australian College of General Practitioners.

The third paper was taken by over 800 trainees around the UK as part of a nationwide study. The results for the trainees in the West of Scotland region are given here but the data from all 800 were used for statistical analysis. The reasons for using a variety of sources are discussed later.

Predetermined pass rates were 68% (RCGP), 70 (±3)% (locally constructed paper) and 76 (±2)% (RACGP). None of the trainees fell below the predetermined pass mark for the RCGP paper. For the locally constructed paper, the mean score was 79.25%, with a standard deviation of 5.46. Five trainees scored below 70%, with marks of 68, 69 (two trainees each) and 64% (one trainee). This one trainee can be reliably identified as falling below the pass mark (95% confidence). The mean score for the RACGP paper was 75%, with a standard deviation of 4.23%. Out of the trainees taking this paper, 57 fell below the pass mark and 20 outside the 95% confidence interval. The Alpha coefficient of the locally written paper was calculated to be 0.76, and for the paper based on the RACGP examination 0.72.

It was clear at this stage that the pass mark set for this third paper produced a much higher fail rate than the others. In view of the fact that more than half of the trainees taking this paper failed to reach the pass mark, it was considered inappropriate to use this pass mark — therefore, using the Hofstee technique,[8] the pass mark was adjusted to 70%, corresponding to the pass mark in the earlier papers. Using this, no trainee fell below the 95% confidence level.

Audit

A total of 345 audits were received, of which four were rated unsatisfactory using the criteria listed. These trainees were given feedback and invited to resubmit. None of the trainees were failed as a result of the audit. In the first year of the study, 14 trainees did not submit an audit.

Videotape assessment

Videotapes were received from 358 trainees: one in the first year of the study refused to submit a tape. Each videotape, of at least

six consultations, was viewed independently by two assessors. The assessors were asked to identify for further assessment those trainees about whose clinical competence they had doubts. Full details of this method are described in a previous paper.[4] The assessors rated 72 trainees as 'refer'. All of the non-referred tapes were reviewed in an attempt to identify any missed referrals and one was identified as such.

One trainee was identified as of dubious competence to practice by both the multiple-choice paper and the videotape analysis.

Overall results

There was some correlation among the four components, but most of the trainees who were identified as of questionable competence were picked up on only one component.

A total of 77 trainees (22%) were identified as being of potentially doubtful competence and were entered into the next level of the process (Table 2); however, 15 trainers exercised their right to halt the process at this stage and issue pass certificates. The three trainees referred by their trainers were refused certificates after discussion at regional level. The remaining 59 were reviewed by the authors, and of these, 22 were referred to external assessors, who concluded that 11 were of acceptable competence and 11 were not. In an attempt to evaluate the possible results for the 15 trainees who did not continue to the next phase of the process, the videotapes of those trainees where both video assessors had felt that the trainee should be referred were examined. All of these tapes were screened by one of us (LMC) and four tapes, which appeared to be of unacceptable standard, were reviewed by 22 of the panel of assessors using a validated instrument.[4] The panel agreed that there was considerable doubt concerning the competence of three of these trainees.

From these results, it appears that 17 trainees would have been refused certificates if the results had been mandatory. However, as stated above, the decision to issue a certificate remained the trianer's perorgative at this stage. This would represent 5% of trainees. In fact, only seven trainees were refused certificates (three as result of the videotape, three as a result of the trainer's report, and one as assessed by both video and multiple-choice test). The other 10 trainees were issued certificates by their trainers.

Discussion

Although not all of the trainees took part in the whole process, more than 95% completed all components. No hard data are available on the non-responders but all received certificates of satisfactory completion. The external assessors agreed that there was legitimate cause for further analysis of all of the trainees referred to them. Analysis of the videotapes of the trainees who were not referred identified one trainee of doubtful competence who had been missed in the initial assessment.

Although there was some correlation among the four components they clearly identified different problems:

- the trainer's report identified trainees who were disorganized, unreliable or unable to work at an acceptable pace
- the videotape picked up problems in communication and patient management skills
- the audit identified trainees who could not grasp the concepts of carrying out quality assurance or who were unable to describe it
- the multiple-choice paper pointed out deficiencies in knowledge.

Table 2. Overall results of summative assessment of 359 trainees.*

Assessment component	Number submitted	Number referred	Number unacceptable
Trainer's judgement	359	3	3
Audit	345	4	0
Video assessment	358	72	14
Multiple choice paper	358	1	1
Total number of trainees		77	17

*One trainee who failed the multiple-choice paper also failed the video, two trainees who were referred as a result of the audit were also referred by the video.

Individual components

Multiple-choice test. The first two multiple-choice papers used in the assessment produced a small number of unsatisfactory scores. If the scores within one standard error of the pass mark are excluded, the performance of only one trainee was unacceptable. It is planned to further modify this paper by introducing a larger number of problem-solving and decision-making tests. By setting the pass mark using the Angof technique, a low or non-existent failure rate was obtained on two occasions, and on a third occasion, using a panel of experienced regional and associate advisers, it produced a minimum acceptable mark which turned out to be higher than the mean mark obtained by the candidates. A refinement of this technique — the Hofstee process — may be the best way forward. In this situation, the panel, as well as performing the Angof assessment, decides acceptable maximum failure rates. This system could be construed as a return to peer referencing but it is important to try to retain the possibility that all candidates could pass. It would be simple to fail the bottom 5% or so but this would be unfair, at least in theory, and probably politically unacceptable. It is perfectly possible to envisage a cohort of trainees in which every individual was of acceptable competence. Therefore, a system for assessment of minimum competence must be able to deal with such a situation. The trainee subcommittee of the General Medical Services Committee would certainly object to such an arbitary cut-off point.

Audit. None of the audits submitted resulted in refusal of a certificate, although a number of trainees did not submit an audit and some audits (even after resubmission) were still judged unsatisfactory. However, it was considered at this stage that refusal of a certificate would be unjustified until further validation had been performed.

Trainer's report. The trainer's judgement produced three referrals for further assessment. The UK advisers intend to use a trainer's report with considerably more questions than used in this study. This will encourage the trainer to think more carefully about the trainee's competence and may produce more referrals.

Videotape. The video component produced most referrals. As discussed in earlier papers, it was anticipated that about 20% of trainees would be referred and that 2–5% would cause serious concern. This was confirmed in the full-scale study.

The relatively large number of trainees initially identified by the video as unsatisfactory who were subsequently judged to be satisfactory is unavoidable if the unsatisfactory doctors are to be identified. As in many situations, an increase in specificity could be achieved only by a reduction in sensitivity. It would be tempting to consider for referral only those trainees identified by both

assessors; however, in one such case, the panel of assessors agreed that the trainee's performance was in fact unsatisfactory.

Credibility of summative assessment

An important issue is the trainer's ability to overrule the findings of the summative assessment process. Such assessment will never be credible while the trainer retains the right to issue a certificate regardless of the results of the assessment. The unshakeable belief of some trainers that their trainee was competent, despite evidence to the contrary, was striking — it is interesting that the trainer was unable to produce any objective evidence of competence in any of these cases.

One of the major questions concerning summative assessment is the opportunity cost, i.e. are the results worth the time and money expended, or could resources be better used elsewhere? Is it economical to devote significant time and resources to a process which will result in relatively few trainees being refused a certificate? A detailed analysis of the resource implications of summative assessment (UK Conference of Regional Advisers, unpublished data) has been carried out, but is outwith the scope of this paper. For the video component, which is the most resource-intensive, 2 h of assessor time is required per trainee for initial assessment. The national panel would look at 10% of trainee videotapes: one day's work for 40 assessors each year.

In this study, the assessment process resulted directly in seven trainees being refused a certificate, four of whom had been deemed acceptable by their trainers. A further 10 might have been refused if the process had been mandatory. This is a considerable increase on current numbers. How far we can extrapolate these figures to the UK as a whole is unclear: it is possible that the proportion of unsatisfactory trainees in the West of Scotland region is different from that of the rest of the country but elucidation of this will be possible only when other regions have taken part in the process.

It has been argued that improving the system of formative assessment would render summative assessment redundant.[7] Although we agree with this in principle, the West of Scotland results were obtained in the presence of an extensive and mandatory formative assessment programme.[9]

Conclusions

The West of Scotland summative assessment programme identified a number of trainees who were not competent to enter general practice. Most of the unsatisfactory trainees were identified by observing their behaviour in consultation. When this assessment process becomes mandatory, approximately 5% of trainees are likely to be refused a certificate. This has resource implications for further training of these doctors. The results of this study also have considerable implications for patients, those involved in training and government. Current vocational training regulations must be altered for summative assessment to be effective.

References

1. Joint Committee for Postgraduate Training for General Practice. *Assessment working party report.* London: Joint Committee for Postgraduate Training for General Practice, 1993.
2. Campbell LM, Howie JGR, Murray TS. Summative assessment: a pilot project in the West of Scotland. *Br J Gen Pract* 1993; **43:** 430–434.
3. Chairman's statement, UK Conference of Regional Advisers, Cambridge, 1994.
4. Campbell LM, Howie JGR, Murray TSM. The use of videotaped consultations in the assessment of general practice trainees. *Br J Gen Pract* 1995; **45:** 137–141.
5. Livingstone SA, Zieky MJ. *Passing scores.* Princeton, NJ: Education Testing Service, 1982
6. Irvine D, Irvine S (eds). *Making sense of audit.* Oxford: Radcliffe Medical Press, 1991.
7. De Gruijter DNM. Compromise models for establishing examination standards. *J Educ Measurement* 1985; **22:** 263–266.
8. Banks I. English National Trainee Conference, 1994
9. Campbell LM, Murray TS. Trainee assessment – a regional survey. *Br J Gen Pract* 1990; **40:** 507–509.

Address for correspondence
Dr L M Campbell, Department of Postgraduate Medical Education, University of Glasgow, Glasgow G12 8QQ.