Summative assessment: a pilot project in the west of Scotland

L MALCOLM CAMPBELL

J G R HOWIE

T STUART MURRAY

SUMMARY. In the autumn of 1991 the Committee in General Practice of the west of Scotland region appointed a working party to investigate the possibility of developing a credible, valid and reliable programme of summative assessment for general practitioner trainees. The working group formulated a four-part package consisting of a multiple true–false paper, a trainee audit project, the trainers' judgement, and analysis of videotaped consultations. The reasons for the use of this selection of methods are discussed. It is suggested that a summative assessment process for trainees should make use of the trainers' considerable knowledge of the trainee, have an external component, be criterion referenced, have an element of continuous assessment, and involve direct assessment of clinical competence. A pilot study of assessment of clinical competence using videotapes of routine trainee consultations by 25 volunteer general practitioner assessors is described. A rating instrument for use in differentiating the competent from the not yet competent trainee is discussed.

The working group and the group of videotape assessors came to the provisional conclusion that the use of videotaped consultations may be a valid and feasible method of assessing the competence of general practitioner trainees as part of a balanced summative assessment programme.

Keywords: vocational training assessment; educational assessment; assessment techniques; videotape recordings.

Introduction

CONSIDERABLE work has been carried out on the assessment of trainees in general practice. Mulholland and Tombleson have produced a useful theoretical analysis of summative assessment.[1] In 1990, the chairmen of the General Medical Services Committee, Royal College of General Practitioners and Joint Committee on Postgraduate Training for General Practice stated that the certificate of satisfactory completion of vocational training was in fact a certificate of competence, as opposed to a certificate of attendance.[2]

A survey of general practitioners carried out by the General Medical Services Committee showed that 47% of trainers agreed with the statement that 'the vocational training certificate issued by the JCPTGP provides sufficient proof of a GP trainee's competence to practise as a GP', while 44% disagreed with this statement.[3]

In view of these developments, in the autumn of 1991 the Committee in General Practice of the west of Scotland region set up a working group to explore the possibility of developing a credible, valid and reliable method of summative trainee assessment. The working group consisted of the postgraduate dean, regional adviser, assistant adviser, two representatives each from the west of Scotland faculty of the RCGP and the Scottish General Medical Services Committee and two trainers from the west of Scotland. This paper describes the work of that group to date in formulating the overall structure of the assessment package and in particular in the use of videotaped consultations in summative assessment. Although there is as yet no evidence concerning the validity and reliability of the process it is important to document progress made, particularly since the working party on summative assessment of the Joint Committee on Postgraduate Training for General Practice has drawn attention to the pilot scheme in its report (unpublished, 1993).

Pre-existing regional assessment programme

The west of Scotland region has on average 155 trainers with 150 trainees in post at any one time. A survey of local assessment procedures was carried out in 1989.[4] This demonstrated that a wide range of assessment methods were being used formatively in the region but that 10% of trainees were receiving virtually no assessment within the practice. As a result of this survey a regional formative assessment package was introduced in the west of Scotland in July 1990:

August	Confidence check list, multiple choice paper.
September	Formative video assessment.
October	Trainee interview with associate adviser.
November	Manchester rating scales,[5] objective structured clinical examination.[6]
January	Formative video assessment.
February	Check list, multiple choice paper.
March	Manchester rating scales, objective structured clinical examination.
April	Formative video assessment.

This form of assessment is now a mandatory minimum for all training practices in the region. The purpose of this programme is to provide trainers with the evidence they need in order to determine the educational attainments and needs of the trainee, that is, formative assessment.

Attributes of summative assessment system

In order to devise an appropriate system of summative assessment it is necessary to define the attributes of the ideal system. The following is the list of proposed attributes of a summative assessment programme agreed by the working group:

- Trainer's assessment should carry weight.
- There must be an objective external contribution.
- Clinical competence must be directly assessed.
- Performance throughout the trainee year should count in the assessment.
- A 100% pass rate should be possible.
- The procedure must be feasible.

The trainer has the opportunity to form a judgement based on the trainee's performance over the entire period of training. However, because of the close and friendly relationship which almost always develops during the trainee period it becomes difficult for

L M Campbell, FRCGP, assistant regional adviser (assessment) and T S Murray, PhD, FRCP, FRCGP, regional adviser, West of Scotland Committee for Postgraduate Medical Education, University of Glasgow. J G R Howie, MD, PhD, FRCPE, FRCGP, professor of general practice, University of Edinburgh.

Submitted: 26 October 1992; accepted: 9 February 1993.

the trainer to take the decision that a trainee is not competent for independent practice. The possible conflict of interest is such that an external contribution to the assessment process is highly desirable for the credibility of the system, in the same way as all universities have external examiners in degree examinations. For any assessment method to attain face validity it must measure an area which is relevant to the eventual professional activities of the candidate. The ability of the doctor to carry out consultations successfully is therefore a major determinant of the doctor's overall competence. It has been shown, for example, that general practitioner trainees with good interviewing skills are more likely to offer relevant advice and treatment to patients with psychiatric disorders.[7]

Indirect methods of assessing clinical ability can be used but there is as yet no evidence for a close correlation between these methods and actual clinical competence. There is in fact some evidence that multiple choice papers and modified essay papers are not by themselves good predictors of postgraduate performance.[8] In addition, multiple choice papers may affect the student's approach to learning, producing a superficial rote learning approach.[9] Individuals have a varying response to the stress of an end-point examination. Some element of continuous assessment would therefore appear to be desirable. It was felt inappropriate to have a pre-determined failure rate as occurs in many postgraduate examinations. However, the use of criterion referencing rather than peer referencing is then necessary where possible. There are obvious difficulties in defining criteria for competence in general practice; the two possible approaches are to produce a detailed list of attributes or to define competence in broad but imprecise terms. The working group took the view that detailed criteria would be impossible to produce and the latter course was therefore chosen. There is some evidence that such global scales are at least as reliable as more complex marking systems,[10] although it must be acknowleged that broad evaluations can come close to peer referencing since the performance of the trainee is being compared with that of a 'standard' trainee.

Legal position

At present the general practitioner trainer and relevant hospital consultants have the statutory responsibility to sign the certificate of completion. The working group took the view that there need be no change in this system — it was felt that the summative assessment results should guide the trainer in making the decision to sign. Where a trainer's views are consistently in disagreement with the other findings of the assessment process this could become an issue in the regular reapproval procedure of the training practice. Clearly this has an air of coercion about it but it is the duty of regions to satisfy the Joint Committee on Postgraduate Training for General Practice that standards of assessment are adequate. Facilities would be necessary for the further training of the small number of trainees identified as not competent by the assessment process. At the moment a further six months training is available for trainees who appeal unsuccessfully against the refusal of a certificate.

Options for summative assessment

The working group identified three options for a summative assessment system: maintaining the *status quo*, using the membership examination of the Royal College of General Practitioners or devising a new system.

As regards maintaining the *status quo,* in the west of Scotland region the granting of a certificate is largely based on the trainer's judgement, informed by the current formative assessment schedule. However, it was felt that this approach would lack credibility with the public and regulatory authorities because of the absence of any contribution from outside the practice. Further, to ask trainers to be teachers, mentors, and then become pass/fail assessors would destroy the relationship which should exist between trainer and trainee.

The RCGP membership examination has been developed over a period of years and its reliability has been extensively studied.[11] However, as presently constituted the examination has two main drawbacks. First, there is no direct assessment of clinical performance; this was noted as long ago as 1979.[12] Secondly, the examination is peer referenced in that the pass rate is effectively fixed at around 75%. An additional problem is the lack of trainer input.

The working group therefore decided to attempt to devise a new system for the region which had all of the attributes listed above.

Methods used in summative assessment

The working group decided to use four components in the assessment process; a multiple true–false paper, trainers' overall judgement, a completed audit carried out by the trainee and an assessment of videotaped consultations.

The group felt that each of the four components had specific features which would combine to produce a balanced overall assessment. Factual knowledge is obviously important in general practice. A properly constructed multiple choice paper is a reliable and feasible method of identifying factual knowledge. One region in England requires a minimum acceptable performance in a multiple choice paper before a potential trainee will be accepted into a practice.[13] As stated earlier the aim was to use criterion referencing where possible. A multiple true–false paper is effectively an instrument for ranking trainees, but by setting minimum levels of performance it was hoped to avoid a pre-set rate. The trainer is uniquely placed to observe the trainee over the course of the year particularly in the areas of attitudes and behaviour. However, it was acknowledged that the trainer's judgement will be based to some extent on his or her previous experience of trainees and to that extent will contain elements of peer referencing. Performance review has long been recognized as a necessary component of the practice of medicine — a completed audit will demonstrate that the trainee has absorbed the principles of audit and carried out the practice of performance review. The audit will be judged using pre-defined criteria. The major part of general practice takes place in the consultation with an individual patient. For any system of assessment to be credible it must address this area. Three possible approaches to assessing performance in the consultation were identified: direct observation of consultations, the use of actors in a simulated surgery and video recordings of real consultations.

Direct observation by an assessor has several attractions. The assessor can attempt to confirm the trainee's findings by taking an additional history and examining the patient, discussing the trainee's actions with the trainee and obtaining immediate feedback from the patient. However, the group felt that this approach would have logistic problems in that if there were 25 assessors, each assessor would have to spend a session with each of 10 trainees (this assumes that the maximum number of trainees finishing at any one time (July) is 125 and that for reliability reasons there would be two assessors per trainee; owing to the size of the region this would also involve considerable travelling time). In addition, there are potential problems of disruption of the consultation by the observer. Finally, there would be no possibility of assessing the reliability of the assessment unless the consultations were recorded in addition to being observed.

Much work has been done on the use of simulated patients in the objective structured clinical examination.[6] The examination has been used on a region-wide basis both in the west of Scotland and elsewhere.[14] It provides the opportunity to present the trainees with standardized presentations, but the time

involved would be considerable. For example, to carry out a simulated surgery for 150 trainees with six patients per trainee would require 300 assessors (assuming double marking was used) and say 60 patients with each patient performing 15 times. Fewer assessors and actors could be used by spreading the process over several days but the overall time involved would remain the same and there would be additional problems of contamination. Continuity of care is a major concern in general practice and the objective structured clinical examination with its procession of new patients may be of little value.

The use of videotaped consultations has potential advantages. It enables observation of real consultations in a relatively unobtrusive way. Tapes can be assessed by a number of assessors, thereby measuring reliability and also calibrating the assessors. The group decided that the number of videotape assessors should be small to reduce variability, but large enough to keep the workload to manageable proportions. It was decided to invite applications, initially to carry out videotape assessments but also to develop suitable multiple choice papers and assessment scales for audit, from associate advisers (the Scottish equivalent of course organizers), members of training practices and others with suitable assessment experience — 25 assessors were appointed and no applicants were rejected. The majority (14) were trainers, four were associate advisers and seven were RCGP examiners.

A trainee who was rated as satisfactory in all four components would automatically receive a certificate of completion. A trainee who failed in any one area would enter a referral process. In this referral process the trainee's competence would be discussed by the regional adviser with the associate adviser and the trainer. A review of the trainee's portfolio would be undertaken by a further two assessors, at least one of whom would be from outside the region. If at the end of this process the trainee was deemed not to be competent, refusal of the certificate of satisfactory completion and suitable additional training would be recommended. No clear consensus emerged as to how many trainees might be expected to fail to obtain a certificate under the new scheme but it was felt that it would be unlikely to exceed 5%.

It was agreed that initial effort would be devoted to the assessment of videotaped consultation since the use of multiple choice papers and audit appeared to be straightforward and the trainer's judgement already forms part of the existing formative assessment system.

Assessment of videotaped consultations

A total of 150 trainees were invited by letter in February 1992 to produce four-hour videotapes with accompanying log-books. Trainees were advised to obtain appropriate signed consent from patients and to record all consultations where consent had been given until they had four hours of consultations. Trainees were offered feedback and the return of their tapes if wished. Trainers were informed about the study by the local associate advisers and urged to render all possible encouragement to the trainees. Within six weeks 80 videotapes and log-books had been received at the regional office. Each assessor was sent two tapes for viewing prior to an assessors workshop in June 1992. Thus, the tapes of 50 trainees had been viewed prior to the assessors workshop. The assessors were asked to assess as many consultations from each tape as they felt necessary to come to a firm conclusion about a trainee, that is definitely competent or should enter the referral process, but in any event to view at least six consultations. Each assessor was also sent the trainee log-book and a marking schedule. The assessors workshop was used to identify strengths and weaknesses in the log and marking schedule and to look at tapes where assessors had been doubtful about trainee competence. The main conclusions of the workshop were as follows.

Videotape length and content

In none of the tapes viewed was a length of four hours felt to be necessary to reach a decision about a trainee. A two hour tape including at least two consultations with children and two with patients with chronic problems was felt to be adequate but further analysis of this continues.

Trainee videotape log

The original version of the videotape log for each consultation did not ask the trainee to rate the degree of difficulty of the consultation, nor was the trainee specifically requested to write comments. The workshop felt that by asking the trainee to rate the consultation for difficulty and to comment on it the assessors would find it easier to rate the trainee's performance. There was a clear consensus that a poor consultation where the trainee was aware of the problems would be of less importance than where the trainee had not noticed any problems. The videotape log now includes the following items: trainee name; consultation number; camera clock time; reason for patient's attendance; physical findings if any; action taken, for example prescription issued; and degree of difficulty of consultation (easy/moderate/difficult). The trainee is also asked to write up to 100 words on the setting of the consultation, what was achieved and what issues may arise as a result of the consultation.

Assessment form

Assessors were asked to look at how well the trainee succeeded in carrying out the tasks of the consultation. The tasks chosen were modified from those described by Pendleton and colleagues.[15] The working group chose to use consultation tasks rather than consultation skills as the criteria for assessment. This was because more detailed rating scales which cover skills such as the Hays scale[16] are more complicated to complete and in the sample considered not all of the areas seemed relevant. In addition, the group took the view that outcome was what mattered — a doctor's consultation style should not be a factor provided the appropriate objectives are attained. Clearly a doctor who uses appropriate consultation skills is more likely to succeed in achieving the required results than one who does not and therefore similar results would be obtained from both sets of criteria in most cases.

The list of tasks as modified by the assessment workshop after the experience of assessing tapes is as shown below. The item recording the presence of an error was not part of the original scale but the workshop felt that the addition of this item would enable them to record an error and then concentrate on the rest of the consultation.

- Was there any obvious diagnostic or management error?
- How well did the doctor discover the reasons for the patient's attendance?
- How clearly did the doctor define the clinical problem?
- How well did the doctor tailor the explanation to the needs of the patient?
- How well did the doctor manage the clinical problem?
- How effectively did the doctor use resources of time, investigations and manpower?
- How effectively did the doctor relate to the patient?

The assessors scored the first component, presence of error, as present or not present. The other components were scored on a scale of one to six as follows: 1, definitely refer; 2, probably refer; 3, bare pass; 4, competent; 5, good; 6, excellent. The assessor would then make a judgement of pass or refer on the consul-

tation overall. After viewing at least six consultations the assessor would then make a final judgement of refer or pass. The overall judgement would be influenced by the scores in the individual components but would not be achieved by an averaging process.

Assessment reliability

Assessors' opinions were shown to be reasonably consistent when a selection of tapes were reviewed by the whole group at the assessors workshop. Overall, 45 tapes were rated by individual assessors as showing acceptable levels of competence, one was felt to show that the trainee was not competent and this view was agreed by the workshop, and four were identified by individual assessors as showing a doubtful standard. A detailed study looking at interobserver reliability is under way. The decision was made to keep the number of assessors as low as possible, compatible with workload, in order to minimize inter-observer variability and to give each observer sufficient videotapes to facilitate self-calibration. For example if 25 assessors viewed 125 two-hour tapes in one month this would be expected to involve 10 hours of work for each assessor.

Problems encountered

Some trainees submitted small VHS cassettes. Although only accounting for 2–3% of videotapes these required a disproportionate amount of time to transfer to standard VHS format. The picture and sound quality was reasonable overall but it was clear that some standard advice on these aspects should be produced in any final instruction sheet. In particular, the use of camera clocks and desk top microphones made the videotapes easier to assess.

The working group received 12 letters from trainers and two letters from groups of trainees. Some doctors were concerned about confidentiality and felt that videotapes should not leave the practice. Others were concerned that the tapes could be viewed by non-medically qualified personnel. These problems had been anticipated. All tapes were transferred by hand or using recorded delivery to minimize the risk of tapes going astray. The assessment workshop used the services of a non-medically qualified educationalist who was accustomed to dealing with sensitive material and followed the same confidentiality code as doctors.

The working group reminded the practices in the instruction sheet that appropriate informed consent from patients was required. As a result of requests from practices a consent form was produced. This form has been submitted to the defence societies and the West of Scotland General Practice Ethical Committee whose initial response has been favourable. A guide to trainers on informed consent and an explanatory leaflet for patients will be produced if the scheme goes forward on a universal basis.

Concern was expressed by some trainers that the presence of the camera would disrupt the consultation and prevent the patient discussing problems. A study by Martin and Martin showed that 78% of patients who agreed to be video recorded forgot about the camera during the consultation.[17]

Some trainers felt that they had been inadequately involved in the process leading up to the pilot study. There is no doubt that fuller consultation would have reduced the amount of disquiet, but the working group decided that full consultation was not possible in view of the speed with which the Joint Committee on Postgraduate Training for General Practice was moving forward on summative assessment.

Preliminary results

The panel of assessors were in no doubt that viewing consultations was a powerful tool in assessing competence. Those members of the panel who were RCGP examiners felt that the video recording added an additional dimension to their ability to identify trainees' abilities. The RCGP examiners felt that the ability to watch trainees consult went a long way towards the goal of assessing performance rather than competence. Some difficulties were, however, encountered which indicated that videotapes of consultations cannot provide all the answers. Originally health promotion had been included as one of the parameters for assessment but in many cases it was found to be impossible to decide whether or not appropriate health promotion had been offered, as this depends not only on the nature of the presenting problem but also on data already in the patient's records. For example, questions about smoking would not be relevant if the patient was known to be a non-smoker. Data from the records were unavailable to the assessors. Occasionally the assessors had difficulty where they felt the candidate had made an incorrect diagnosis but to be certain would have had to examine the patient, which was not possible.

The videotaped consultations provided information in unexpected areas. The degree of organization of the individual trainee and the practice was often apparent. A number of trainees did not refer to the patient's notes prior to the consultation, although they may have gone through all of the records prior to the session, and some did not refer to the notes during the consultation. Trainees were heard to ask patients their identity, and to ask questions which, as became apparent later in the consultation, were already answered in the records. More worryingly, some patients had to point out information to the doctor which the doctor needed to know, such as a recent abnormal smear result.

Future plans

A formal analysis is now under way to assess the reliability of the videotape assessment using two assessors per tape. An attempt will also be made to determine the number of consultations which need to be viewed by asking the assessors to rate the degree of impact of succeeding consultations. The working group also hopes to identify which of the components of the consultation in the rating scale are marked most and least consistently. Further modifications may then be made.

In addition, possible correlations between performance as recorded during the consultation and knowledge assessed using multiple choice papers and performance in the RCGP membership examination, will be examined.

The working party of the Joint Committee on Postgraduate Training for General Practice had called for feasibility studies in summative assessment in its report (unpublished, 1993). In the event that summative assessment becomes mandatory the working group hopes to have a package available in a validated form to ensure that any developments are instituted by the profession rather than imposed by the regulatory authorities.

References

1. Mulholland H, Tombleson PMJ. Assessment of the general practitioner. *Br J Gen Pract* 1990; **40:** 252-254.
2. Irvine DH, Gray DJP, Bogle IG. Vocational training: the meaning of 'satisfactory completion' [letter]. *Br J Gen Pract* 1990; **40:** 434.
3. General Medical Services Committee. *Your choices for the future.* London: GMSC, 1992: 101.
4. Campbell LM, Murray TS. Trainee assessment — a regional survey. *Br J Gen Pract* 1990; **40:** 507-509.
5. Centre for Primary Care Research, Department of General Practice, University of Manchester. *Rating scales for vocational training in general practice. Occasional paper 40.* London: Royal College of General Practitioners, 1988.
6. Harden RM, Gleeson FA. Assessment of medical competence using an objective structured clinical examination. *Med Educ* 1979; **13:** 39-54.
7. Millar T, Goldberg DP. Link between the ability to detect and manage emotional disorders: a study of general practitioner trainees. *Br J Gen Pract* 1991; **41:** 357-359.

8. Rabinowitcz HK. The modified essay question: an evaluation of its use in a family medicine clerkship. *Med Educ* 1987; **21:** 114-118.
9. Marton F, Saltor R. Symposium — learning processes and strategies. *Br J Educ Psychol* 1976; **46:** 115-127.
10. Gorman TP, White J, Orchard G. *Language performance in schools: secondary survey report 2.* London: HMSO, 1987.
11. Godlee F. MRCGP: examining the exam. *BMJ* 1991; **303:** 235-238.
12. Hannay DR. Postgraduate examinations for general practice in Canada and the United Kingdom. *Lancet* 1979; **2:** 28-30.
13. Roberts J, Freeman J. Entry to general practice training. *BMJ* 1983; **287:** 1677-1679.
14. Walker R, Walker B. Use of the objective structured clinical examination for assessment of vocational trainees for general practice. *J R Coll Gen Pract* 1987; **37:** 123-124.
15. Pendelton D, Schofield T, Tate P, Havelock PP. *The consultation: an approach to learning and teaching.* Oxford University Press,1984.
16. Hays RB. Assessment of general practice consultations: content validity of a rating scale. *Med Educ* 1990; **24:** 110-116.
17. Martin E, Martin PML. The reactions of patients to a video camera in the consulting room. *J R Coll Gen Pract* 1985; **34:** 607-610.

Acknowledgements

This project is supported by a grant from the scientific foundation board of the Royal College of General Practitioners. The authors would like to acknowledge the help and support given by several individuals connected with the RCGP, in particular Professor Lesley Southgate, convener of the panel of examiners, and Peter Tait for his help in developing the rating scales. We also thank our panel of assessors for giving so much of their time, and the trainees of the west of Scotland.

Address for correspondence

Dr L M Campbell, West of Scotland Committee for Postgraduate Medical Education, University of Glasgow, Glasgow G12 8QQ.