



<https://theses.gla.ac.uk/>

Theses Digitisation:

<https://www.gla.ac.uk/myglasgow/research/enlighten/theses/digitisation/>

This is a digitised version of the original print thesis.

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>  
[research-enlighten@glasgow.ac.uk](mailto:research-enlighten@glasgow.ac.uk)

The development of a summative assessment system for  
Vocational Trainees in General Practice

L Malcolm Campbell

MB ChB (Glasgow) DCH, FRCGP

Submitted for the degree of

Doctor of Medicine

University of Glasgow

1996

ProQuest Number: 10390922

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10390922

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346

GLASGOW UNIVERSITY  
LIBRARY

11129 (copy 2)

GLASGOW  
UNIVERSITY  
LIBRARY

## INDEX

	<b>Page</b>
Title	1
Index	2
Declaration	4
Abstract	5
Introduction	12
Hypothesis	31
<b>Chapter 1</b>	<b>33</b>
An analysis of Trainee assessment prior to the introduction of summative assessment	
<b>Chapter 2</b>	<b>50</b>
Summative assessment : A pilot project in the West of Scotland	
<b>Chapter 3</b>	<b>77</b>
The use of videotaped consultations in the summative assessment of vocational trainees in general practice	
<b>Chapter 4</b>	<b>100</b>
Summative assessment of vocational trainees: The results of the first three years	

<b>Chapter 5</b>	119
The effect of videotaping consultations on patient satisfaction	
<b>Chapter 6</b>	126
Continuous assessment in the Practice after the introduction of mandatory summative assessment.	
<b>Chapter 7</b>	144
The assessment of communication of medical undergraduates in United Kingdom Universities	
<b>Chapter 8</b>	152
General discussion	152
Acknowledgements	168
List of graphs and tables	169
List of Appendices	170
Appendices	171
Bibliography	201

## **Declaration**

In accordance with the regulations of the University of Glasgow, I declare that this thesis has been composed by me. The work reported in this thesis is my own with all contributions from other workers clearly indicated in the text or the acknowledgements section.

L. Malcolm Campbell

## ABSTRACT

Vocational training for general practice has been a legal requirement in the United Kingdom since 1979. It was agreed by the relevant bodies in 1990 that trainers should only issue certificates of satisfactory completion to trainees deemed to be competent. A literature review revealed few studies looking at the effectiveness of this system of certification.

The aim of this thesis was to evaluate the current system of summative assessment and to investigate the possibility of developing a system with increased objectivity, validity and reliability.

The specific objectives were:

- (i) to assess the nature and extent of assessment of trainees during their general practice year
  
- (ii) to explore the possibility of introducing an external and reliable method of trainee assessment
  
- (iii) to carry out an evaluation of the chosen method of consulting assessment

(iv) to identify the effect of the introduction of an external system of summative assessment on the numbers of trainees identified as being not yet competent

(v) to measure the effect of videotaping of consultations on patients

(vi) to evaluate the impact of the introduction of summative assessment on the quality and quantity of practice based formative assessment

(vii) in view of the significant number of trainees found to be below satisfactory levels in consulting skills, to evaluate the nature and extent of communication skills teaching and assessment in the undergraduate teaching of UK medical schools

The methods used to achieve these objectives included retrospective questionnaires and prospective studies of the various assessment tools.

(1) This survey of 125 trainers and 72 trainees demonstrated that where a variety of assessment techniques were used appropriately the trainees found

these to be of value. However, 14% of trainees were exposed to very little assessment, and only 8% of trainees were aware of the presence of a standard assessment programme within the practice. As a result of this survey a region wide continuous assessment package was introduced.

(ii) In this study a working group was set up within the region to develop a reliable and valid summative assessment process. The method arrived at consisted of four independent components all of which had to be passed. These components consisted of :

- consulting skills assessment by evaluation of videotaped consultations.
- a multiple choice question paper
- the evaluation of an audit submission
- a structured trainer's report

The feasibility of video tape production was confirmed by asking the trainees in post at the time to produce appropriate videotapes. 25 attended a workshop after viewing some of the tapes. As a result of the workshop a marking schedule was constructed and an initial impression was reached that acceptable inter observer reliability could be achieved in terms of deciding

the pass/fail status of the candidates. The workload issues for assessors were examined and the conclusion was reached that the exercise would be feasible.

The assessment working group and the panel of assessors attending the workshop came to the conclusion that the use of videotaped consultations could be a valid and reliable method of assessing the competence of general practice trainees. As a result of this exercise it was decided to implement the system locally on a pilot basis.

(iii) This study used videotaped consultations from 10 trainees which were assessed independently by 25 assessors. A total of 1176 consultation assessments were carried out. The conclusions of this study were that using two assessors per trainee tape would produce a 95% probability of identifying an unsatisfactory trainee, while identifying 20% of satisfactory trainees. It was felt that this was acceptable for a screening process. It was also found that assessors judgements became fixed after watching no more than four consultations.

(iv) In this study 359 trainees took part in the summative assessment process. 77 (22%) were initially identified as being of doubtful competence.

17 (5%) were ultimately adjudged to be not yet competent for independent practice. The video instrument produced the highest pickup rate. It was concluded that the videotape instrument reliably identifies non competent trainees. The proportion of trainees likely to be refused certificates will increase to around 5%.

(v) This study examined the scores for consultation satisfaction of 182 patients who had been videoed and 197 seen by the same group of 18 doctors who had not been videoed. There was no difference in satisfaction with the consultation between the two groups. We concluded that the presence of the video camera did not affect patient satisfaction with the consultation.

(vi) A series of questionnaires was used to determine the amount of assessment taking place within the West of Scotland region and the attitudes of trainers and course organisers to the use of videotaped consultations in assessment. The results were compared with the survey carried out in 1989. There had been a significant increase in the quantity of assessment taking place and 94% of trainees stated that analysis of videotaped consultations was taking place. Video was rated useful by most trainees who used it. 76% of trainees claimed to be receiving less than the minimum amount of

assessment stipulated by the region. Trainers rated lack of time as the most important limiting factor to the use of video. An increase in trainer education and monitoring of trainer performance was proposed.

(vii) A questionnaire was sent to all departments of general practice in UK medical schools. In the questionnaire we asked about the teaching and assessment of communication skills. In most schools the department of general practice was the main provider of such teaching. Estimates of students with communication problems ranged from 1-25% with a mean of 14% in the middle year and 12.5% in final year. Only two schools claimed to assess communication skills in final examinations. 23 schools planned an increase in communication skills teaching and assessment.

In conclusion this thesis has demonstrated that the pre-existing system of trainee assessment was unsatisfactory with 14% of trainees being exposed to less than 2 assessment methods during their year (Chapter 2). After 5 years of a mandatory assessment programme only 24% of trainees had been exposed to the mandatory minimum of assessment and 19% had been videoed less than three times in their trainee year. The pre-existing method identified very few unsatisfactory trainees (0.26% in the UK and less than 1% in the

West of Scotland). A new validated system has now been adopted throughout the UK and is likely to increase considerably the number of trainees being refused certificates of prescribed experience and has already increased the West of Scotland figures to 5%. The new system also has implications for trainers, and potential implications for established principals and other specialities.

## **INTRODUCTION**

There is an assumption by the public, shared to some extent by the medical profession, that doctors entering unrestricted practice, either as General Practitioners (GPs) or Consultants, are of proven competence. In recent years there has been an increasing emphasis on the attainment and maintenance of this professional competence including suggestions that there should be regular recertification for all doctors.

A review group chaired by the Chief Medical Officer was set up at the request of the then Secretary of State to look at the identification of poorly performing doctors. The report of this group (Calman, 1995) felt that there was a significant problem and recommended that for General Practitioners 'systems of objective assessment against a national framework should be introduced as soon as possible'. The review group also recommended that Family Health Service Authorities and Health Boards should have a formal role in maintaining standards of professional practice and performance by GPs.

Scully (1995) in a commentary in the *Lancet* suggested that around 6% of established hospital doctors were performing at an unacceptably low level. He suggested that the solution to poor performance was a system of mandatory recertification. We pointed out (Campbell and Murray, 1995a) in response that without meaningful initial certification there could be no such thing as recertification.

A manifestation of concern regarding continuing competence has been the introduction of the post graduate education allowance (PGEA) for general practitioners (Department of Health, 1990). This allowance requires GPs to complete an average of 5 days approved education each year. However, the system is concerned entirely with attendance at approved courses with no requirement to demonstrate educational gain or change in doctor behaviour as a result of educational activities. The available evidence is that very little change in behaviour results from attendance at PGEA approved educational meetings (Kelly M H, submitted for publication) although needs based interactive learning has been shown to improve skills and performance (Houston et al, 1995; Carney et al, 1995).

A compounding factor is that a significant proportion of educational sessions for GPs are provided by the pharmaceutical industry. Of 22,456 sessions attended by GPs based in the West of Scotland in the year to 31st March 1995 6776 (30%) were organised by the pharmaceutical industry. It may be that such courses are relevant educationally but there must be a suspicion that they are geared more to the needs of the Industry rather than the GPs. The other branches of the profession appear, if anything, to be less advanced than general practice in terms of long term monitoring of performance in that their schemes are voluntary and do not even embody financial sanctions (Kemple, 1995).

While it is clearly essential that professional competence, once attained, should be maintained it is equally vital that entry to the specialist ranks should be conditional on the demonstration of an acceptable level of competence as stated by Carney (1992) in a BMJ editorial. Carney argued that a national standard of entry to general practice would put general practice on the same footing as other specialties and suggested that the existing structure based on trainers' assessments was not always reliable due to the 'halo' effect and the variation among trainers.

He postulated that some trainees would never reach acceptable standards but that the majority need not feel threatened and should feel a sense of achievement. In Carney's view formative and summative assessment should become an integral part of every trainee's education.

It could be argued that the lack of an effective method of monitoring continuing competence would be at least partially offset by a reliable, valid and feasible process of initial accreditation. It has been stated that no single assessment method satisfies these criteria (Maguire, 1989). In this thesis I will describe attempts to develop an instrument for assessing competence for entry to general practice and discuss the related issues of validity and reliability. This work may have relevance for other specialities.

### **The need for a system of summative assessment**

The null hypothesis in this thesis is that the current process by which a doctor becomes accredited as a general practitioner is adequate and that no change is required. At the beginning of this work there was very little evidence concerning the reliability and validity of the pre-existing process, which consisted of certification by effectively a single doctor (the GP Trainer) that

the trainee (GP Registrar) was competent. This system of certification - 'the single informed signature' does not occur in any other branch of medicine.

It is known that the number of trainees refused certificates is currently very low. In the 5 years preceding the present study only four trainees had been refused certificates in the West of Scotland region representing less than 1% of the number completing training. Figures from the JCPTGP show that the figures for the United Kingdom (1989-1992) are even lower at 0.26%. Comparisons with Canada which operates a national end point assessment show that the UK fail rate is much lower than that of Canada. These figures do not in themselves indicate that there is anything wrong with the current system but certainly raise questions about current assessment methods. In addition it has been reported that there have been at least two deaths of patients treated by recently certified ex trainees and two young doctors have been struck of the medical register by the General Medical Council (JCPTGP, 1995). In order to test out the hypothesis that the current system is adequate it was necessary to answer the following questions:

- Does the current system identify trainees who are not yet competent to enter unrestricted practice?

- Would an alternative system produce significantly different results in terms of the number of trainees identified as not yet competent?
- Would an alternative system be more valid, reliable and credible than that currently in place?
- Could an alternative system be operated without a disproportionate and unjustifiable increase in expenditure of time and resources?

### **Historical Perspective**

Up until 1975 there was no requirement other than full registration for doctors entering general practice and it was possible to become an unrestricted principal after the completion of two six months pre-registration posts.

In 1975 a body, the Joint Committee on Postgraduate Training for General Practice (JCPTGP), was set up with responsibility for conferring the right to independent practice. This contained representatives from the General Medical Services Committee and the Royal College of General Practitioners, with nominees from other bodies, such as the Universities. By 1979 doctors had to complete a year as a trainee in general practice to achieve certification. In 1982 the regulations were changed and it became necessary to complete 2 years of approved hospital posts and a trainee year. For each of these posts

the doctor responsible for the trainee had to complete a statement of 'satisfactory completion'. The meaning of the term satisfactory completion was not clear at this time, and there was a suggestion that satisfactory in this case could mean simply completing the appropriate time in the post. The situation was clarified in 1990 by a letter (Irvine et al, 1990) signed by the chairmen of the JCPTGP, the GMSC and the RCGP stating that the doctor should have reached an acceptable standard of competence by the end of training.

### **The assessment of competence**

Although the above statement made it clear that the trainee had to be competent at the end of training two questions were left unanswered, firstly what is an acceptable level of competence, and secondly, how do we measure it? The GMC has defined in general terms what a competent doctor is, but not in sufficient detail to permit accurate assessment. The Royal Colleges in their membership or fellowship exams use a peer referenced approach in which a given percentage of candidates fail the exam. There is little published evidence concerning the relationship between passing the various membership exams and clinical competence. There is also very little information about the reliability of the various exams. A fair assessment

method for assessing minimum acceptable competence should, if at all possible be criterion based rather than competitive, on the principle that since all candidates may be of acceptable competence it should be possible for all to pass. This approach adds another dimension of difficulty to developing an assessment instrument.

Havelock et al (1995) have discussed the general characteristics of a profession and how these should apply to potential entrants to general practice. They emphasise that one of the attributes of a profession is self regulation as a result of which those responsible for educating entrants have to ensure that the entrants develop the necessary competence. A further characteristic of professional performance as described by Havelock is the use of expert judgement which learners have to develop. The expertise of professionals is manifest in characteristic behaviour which any assessment process, to be valid, will have to take into account.

The features which have been consistently identified (Schon, 1983; Benner, 1984; Berliner, 1987) include:

- Selective information gathering

- Prioritising
- Awareness of context and background

### **Competence versus performance**

In the assessment of the clinical behaviour of doctors it is day to day performance which is clearly of most relevance (Rethans et al, 1990) but unfortunately this is very difficult to measure. Any form of observation will have an effect on the behaviour being observed and it has been shown by the use of simulated patients (not known to be simulated by the doctor) that doctors tend to perform at a level which is below that of which they are capable (Pieters et al, 1994). It is easy to see why for reasons of time and lack of interest we tend to take short cuts and perform below optimum levels.

### **Methods of assessing competence**

The type of competence being assessed will vary depending on the medical specialty involved but the required attributes of the assessment process remain the same in all specialties. The method should be reliable, valid and feasible. A reliable method is one which will consistently produce the same results at different times, using different assessment materials and using different assessors where applicable. Validity is probably one of the most

abused words in the language of medical education. Fundamentally it means that the assessment method should be measuring something which is important in terms of the doctor's ability to carry out his job. It has been subdivided into various subtypes which include the following:

*Face validity* - This basically means that the method 'looks' valid but has no other supporting evidence.

*Content validity* - The assessment should include a representative sample of the area under test

*Predictive or outcome validity* - This is the holy grail of assessment and means that the method successfully predicts the future performance of the individual being assessed.

To paraphrase it could be said that reliability is about counting what you can measure whereas validity is about measuring what counts.

Any assessment process must be feasible, not simply in the sense that it can be done since given sufficient time and money almost anything can be done,

but in the sense that it can be done with an input of time and resources which are in some way proportional to the outcomes of the process.

### **The development of valid and reliable assessment methods**

The ideal assessment instrument would predict the level of actual clinical performance. This goal has not yet been reached (Hojat et al, 1992). Considerable work has gone into establishing the validity and reliability of assessment methods, particularly in the United States (Edelstein and Ruder, 1990) where legal challenge can occur but also in Australia (Hays et al, 1990a), Canada, Holland (Pieters et al, 1994) and the UK (Godlee, 1991). There are broadly two types of assessment in use. The first involves dealing in some way with written material, the second in engaging in some form of assessed clinical work.

The traditional viva examination is probably an attempt to look at clinical work in a surrogate fashion without a great deal of reliability (Hubbard et al, 1965). The written material may consist of multiple choice question papers in their various forms, essays of various types or patient management problems.

The assessed clinical work ranges from routine unstructured observation in the clinic or operating theatre, to the use of standardised simulated patients or the structured assessment of doctor patient interactions either in real time or on videotape.

The use of written material is the most convenient in terms of time and resources, which is presumably why it forms the mainstay of most undergraduate and post graduate exams. Maguire (1990) expressed the view that the methods widely used to assess competence are largely outdated, that better methods are available, and that the bodies responsible for both undergraduate and post graduate education need to use them. He postulated that the essential skills required for a doctor are generally agreed and include: establishing a rapport with patients, eliciting accurate information about patients' problems and establishing patients' reactions to them; conducting an examination of physical and mental state; selecting and interpreting investigations; showing diagnostic ability; undertaking education, reassurance and counselling of patients; and managing patients in the immediate and the long term. Maguire suggested that these attributes would be best assessed by using detailed and direct audit of skills. The advantages of such an approach would include the effect on students' learning behaviour and in encouraging

teachers to modify the curriculum so that key skills are adequately covered. He contended that traditional pen and paper assessment methods encourage learners to adopt a surface rather than a deep approach to education and to emphasise facts rather than skills.

Written material is, however, potentially reliable to an extent which is very difficult to attain with the assessment of clinical work. The coefficient alpha is the most used measure of reliability in this kind of material and basically relates to the ability of the test to produce consistent results. As far as individual test items are concerned discrimination is used as a measure of reliability.

This basically describes the ability of the item to discriminate between candidates who do well in the test as a whole and those who do badly. A strongly positive correlation indicates reliability while a negative correlation indicates that good candidates do less well in the particular question than bad candidates. Such questions should not survive for long.

Written material, particularly multiple choice papers, can therefore be made reliable. The question is how valid are they? This is vital since there is no

point in having an extremely reliable assessment tool with no validity. Much work has been done in attempting to look at the reliability of written material. Norcini et al (1985) demonstrated that multiple choice papers are more efficient and reliable than paper based patient management problems. Much less evidence is available in terms of validity (Jones et al, 1990). There appears to be only a modest correlation between performance in written tests and in clinical rating scales (Dowaliby and Andrew, 1976). Dwyer (1988) has stated that paper based measurement of psychosocial attitudes and competence are no substitute for direct assessment. Maguire (1985) in a study of terminal care showed that some experienced doctors lacked key interviewing skills. This lack was only demonstrable by direct assessment.

When we come to look at observation of clinical performance the problems of validity and reliability are reversed. If we watch a doctor at work there is a *prima facie* case that the assessment is valid in the sense that we are looking at real performance rather than a surrogate measure. Clearly we have to be sure that we are measuring adequate and appropriate aspects of performance but provided we get that right we can at least claim face validity.

However, it is here that reliability becomes much more difficult to achieve (Herbers et al, 1989) but there is some evidence that reasonable reliability can be attained (Williams et al, 1987). There are several factors which tend to reduce reliability in tests of clinical competence. One factor is marker variance. Unlike an MCQ answer it is possible for equally skilled markers to form differing judgements. This can be minimised by the use of well defined marking schedules and intensive examiner training. An example of this is the oral examination for membership of the RCGP. A recent paper has described in detail the training of examiners and the development of marking schedules (Wakeford et al, 1995) but unfortunately has produced very little data to support a resulting increase in reliability.

As we wrote in the British Medical Journal (Campbell and Murray, 1995) Wakeford and colleagues describe the work which has gone into developing the vivas for membership of the Royal College of General Practitioners and commend this process to other organisations. However, the main point of improving examiner training and exam structure is to improve the reliability of the exam. The part of the discussion in Wakeford's paper which deals with reliability is brief and confusing. There are two separate vivas each carried out by a different examiner pair. The 94% near agreement quoted by

the authors appears to be between the two examiners in each individual oral although the text is self contradictory on this. There is insufficient analysis in the paper to make the 94% figure mean very much.

Factors which should have been considered include the following. Do markers congregate around the middle of the scale? If this is the case such concordance could explain a lot of the apparent agreement. Markers can and do change their gradings after discussion which clearly makes for agreement. Although examiners do not discuss their views on candidates until after the initial marking it is impossible to sit beside one's co-examiner for 30 minutes without forming a (usually accurate) impression of what the co-examiners mark is likely to be. All of these factors would tend towards increasing inter examiner agreement but not necessarily to increasing the reliability of the vivas. A more thorough analysis would enable us to judge if the commonly held view that vivas are inherently unreliable can now be revised. Wakeford and colleagues accepted our points but stated that they currently do not have any data to carry out a reasonable reliability analysis.

Further problems with the reliability of clinical assessment are the apparent lack of consistency of candidates in performing different tasks (Roberts and

Norman, 1990) and the need for large numbers of test items (Newble and Swanson, 1988) to deal with this.

### **The assessment of minimal acceptable competence for entry into general practice**

In this thesis I shall describe the process leading to the development an assessment package to identify those trainees not yet competent to enter general practice. Although specifically relating to general practice it could serve as a template for other specialties.

Before attempting to decide on specific test instruments it is vital to define the attributes to be tested since it is clearly impossible to decide if a doctor is competent without first deciding what aspects of knowledge, skills and behaviour make up competence. In the area of general practice this is particularly difficult because of the wide ranging nature of the job. Where defined technical skills make up a significant proportion of the doctors work this task should be much easier, for example in the specialties of surgery and radiology.

The JCPTGP set out the following basic attributes required in a doctor at the end of training:

1. Adequate knowledge
2. Adequate problem solving skills
3. Adequate clinical competence
4. Adequate consulting skills
5. Adequate skills in producing a written report of practical work in general practice
6. Adequate performance of skills, attitudes and knowledge.

As can be seen these criteria are not defined in absolute terms and require judgements to be made as to what is adequate in any given situation.

Although, ideally, criteria should be defined in absolute terms our assessment working group took the view that such definitions were impossible in general practice. It is difficult to envisage any branch of medicine where it would be possible to define competence in concrete terms. Some, such as cardio pulmonary resuscitation, can be defined but are in a small minority.

It became clear that no single method would assess these attributes and we proposed a multi-format approach which I will describe in Chapter 2.

Chapter 3 describes the development of an instrument to assess consulting competence and in Chapter 4 I record the results of the first 3 years of using the package. The effect of the video camera on doctors and patients is addressed in Chapter 5. In Chapter 1 I examine the level of assessment carried out in practices prior to the introduction of a formal regional programme. The effect of the introduction of summative assessment on training is discussed in Chapter 6. In Chapter 7 I examine the grounding in consulting skills which our medical students receive. Finally in my conclusions I will discuss the possible future role of the summative assessment package.

## **HYPOTHESIS**

The system of summative assessment of general practice vocational trainees in use at the time of writing of this thesis, the certificate of satisfactory completion, provided a valid and reliable measurement of competence.

Prior to 1979 the only requirement for entry into general practice was full registration with the General Medical Council. Since then it has been a requirement to obtain a certificate of prescribed or equivalent experience from the Joint Committee on Postgraduate Training for General Practice. This certificate is awarded on satisfactory completion of one year as a trainee in an approved training practice and a minimum of two years in approved hospital posts. Satisfactory completion is certified by the issue of certificates by the general practice trainer and the relevant hospital consultants.

This thesis developed as a result of the analysis of continuous assessment we carried out in 1989 (Chapter 1). This showed such a wide variation in the quality and quantity of assessment being carried out in the training practices in a very subjective fashion that it seemed inevitable that not all non-competent trainees would be identified.

As a result of these findings we entered the second phase which consisted of an analysis of the desirable characteristics of assessment of competence. Thereafter a working group was constituted to identify appropriate ways in which to develop more valid and reliable assessment methods (Chapter 2).

The third leg of the study was to develop a consulting assessment tool (Chapter 3) and pilot and modify as appropriate the four chosen components using adequate numbers of subjects (Chapter 4).

Additional issues then arose such as the effect of video recording on patient satisfaction (Chapter 5), the effect on in practice assessment of a summative assessment programme (Chapter 6), the teaching of communication skills in medical schools (Chapter 7) which were investigated. Finally there arose the opportunity to put the system into operation throughout the United Kingdom (Chapter 8).

## **CHAPTER 1**

### **An analysis of trainee assessment prior to the introduction of summative assessment**

#### **Background and Objectives**

The West of Scotland Region is a large area and at the time of this study had on average 140 trainees in post, most of whom start in August. Just under half of the trainees were on three year schemes. Training Practices vary from large inner city Health Centres to single handed isolated rural Practices. There was one Regional Adviser, one Assistant Adviser and 15 Associate Advisers. Most of the Associate Advisers fulfil the equivalent role to that of Course Organiser being responsible for running the 11 different trainee day release programmes in the Region.

For several years the region's priority objective in training had been to satisfy the JCPTGP criteria regarding the structure and organisation of training practices. Considerable work on assessment had been carried out, with particular emphasis on an annual multiple choice paper and training courses for Trainers on teaching on the consultation using the techniques described in

"The Consultation" (Pendleton, 1984). In 1989 it was decided that increasing emphasis would be placed on Trainee Assessment and an Associate Adviser (Assessment) was appointed. It was agreed that the first task would be to measure the use of, and attitudes to, a range of assessment methods. In addition an attempt would be made to gauge the regional attitude to the possible introduction of a standardised assessment programme, and the use of the MRCGP exam as an end point assessment for trainees. In this chapter I shall describe the results of this survey.

### **Method**

In October 1989 a postal questionnaire was sent to the 140 trainers in the region and to the 102 trainees who had completed their trainee year in July 1989.

The questionnaires sent to both groups were identical apart from minor changes in wording designed to make the questions relevant to the two groups. Each questionnaire was accompanied by an explanatory letter which contained an assurance of anonymity.

The first part of the questionnaire asked about the use of six well known assessment tools. These were:

- Videotaped trainee consultations
- Trainer sitting in on trainee surgery
- Manchester rating scales (Freeman 1976, RCGP 1988)
- Topic check-list
- Written work in the Practice
- Objective Structured Clinical Examination(OSCE)

For each of these methods the respondent was asked if the particular method was used in the trainee year, and if so how often. All respondents were then invited to rate the particular method on a five point scale for usefulness in assessment as shown below:

1	2	3	4	5
useless				very useful

The remaining five questions required simple yes/no answers. These were:

- Is hard data used for assessment e.g. prescribing statistics, referral rates?
- Are any other assessment methods used?
- Is there a standard trainee assessment programme in the Practice?
- Should there be a standard region wide assessment programme?
- Is the MRCGP exam a valid assessment at the end of the trainee year?
- In addition respondents were invited to add comments at each question.
- Statistical comparison of groups was carried out using Chi<sup>2</sup> with appropriate degrees of freedom.

## **Results**

Of the 140 trainers in the region 125 (89%) responded to the questionnaire. Of the 102 eligible trainees 61 (60%) responded to the first mailing. This low response rate was at least partly due to the difficulty of tracing the trainees, many of whom, had moved away from the area. A second mailing produced a final response of 72 (70%).

Table 1 shows the recorded use of assessment methods. There is a close correlation between the two groups in terms of the "obvious" assessment methods i.e. Video, OSCE, and Written work. There is a marked difference in the areas of Sitting in, Manchester Ratings, use of check-list, hard data,

and "other" methods. In all cases a higher proportion of trainers than trainees claimed to use the techniques.

Figures 1 and 2 show the number of assessment methods recorded for each group. The mean number of methods used by trainers was 5.2 with a standard deviation of 1.6. For trainees the mean was 3.7 with a standard deviation of 1.8. Two trainees claimed to have had no assessment of any kind while one trainer claimed to have carried out at least 9 forms of assessment. Both sets of results produce a normal distribution curve suggesting a homogeneous group.

Tables 2 and 3 show the trainer and trainee scores for each assessment method together with the ratings for users and non users of each method.

In all cases with the exception of the trainees views on Manchester ratings the mean scores were higher in the group with experience of the technique. It should be noted in the context of Manchester ratings that this is a method for combining and reporting the results of the other methods. It is therefore understandable that the use of Manchester ratings per se are not rated more highly by users than non users.

The scores of users and non users were compared by using the Chi Square test. This test enables comparison of the scoring patterns in a more detailed way than simple comparison of means. In the case of trainers the scores were higher in the group using a particular method. As can be seen this was statistically significant for all methods except the OSCE where significance did not quite reach a 5% P value. For trainees the positive trend in users was significant except for Manchester ratings where there was no difference between the groups. It is interesting that many of the group comparisons reached a significance of  $P < .001$ .

Table 4 shows the responses to the questions on standard assessment programmes, the desirability of a regional programme and the validity of the MRCGP exam.

Further analysis of these groups showed some statistically significant correlations with the answers to other questions. Those trainers who stated that they operated a standard assessment programme were more likely to use a larger number of assessment methods ( $P < .001$ ), to use video ( $P < .025$ ), to rate video highly ( $P < .005$ ), to rate Manchester ratings highly ( $P < .005$ ), to use written work ( $P < .025$ ), and to mention additional assessment methods

( $P < .001$ ). Analysis of those trainers who favoured a regional assessment programme showed a negative correlation with using the check list i.e. those using the check list tended to be against a regional programme and vice versa ( $P < .025$ ). There was no apparent relationship between views on the MRCGP exam and any of the assessment methods.

Analysis of the trainee groups showed there was a correlation between those trainees who had a standard assessment programme and those with a high score over the whole range of procedures ( $P < .01$ ).

### **Discussion**

Considerable work had been done prior to this study on developing assessment methods. In 1976 Freeman and Byrne published the Manchester rating scales and since then many other methods have been developed and used for assessment. The OSCE has gained acceptance because of the perceived inadequacies of rating scales and has been used on a region wide basis for assessment (Walker, 1987). An informal survey of all training regions in the United Kingdom carried out prior to undertaking this study indicated that all of the methods covered in this study were widely promoted at regional level.

However, there was little evidence that trainers were entering enthusiastically into assessment and a study of a small number of trainers indicated that few saw a need for either formative or summative assessment of trainees (Taylor, 1988).

By the time of this survey the JCPTGP had given strong support to the use of a wide variety of assessment methods in the trainee year (JCPTGP, 1987). It is generally agreed that formative assessment is important in the education of trainees. Without such assessment it is impossible to measure the trainees areas of unmet need or the effectiveness of teaching. The Joint Committee for Postgraduate Training in General Practice had at the time of the survey produced no firm guidelines on assessment although a working party report recommended the use of Manchester Ratings (JCPTGP, 1987). This technique was the only one not to be rated more useful by the trainees who used it than those who did not. It seems clear that as yet no single form of assessment can provide all the information needed. It is therefore important to use appropriately a wide range of assessment methods. In this survey the reported use of assessment procedures differed in the trainer and trainee groups. For all methods the trainers reported a higher usage than the trainees.

Because the response rate was not 100% it is possible that the groups were skewed but in fact the figures for such methods as Video and OSCE are very similar. This would suggest that there is a degree of over and under reporting. If we take the trainee figures as being the lower end of the probable range it would seem that 81% of trainees were exposed to between 2 and 6 assessment methods per year while 14% used less than 2. This may underestimate the number of methods used but it could be argued that if the trainees were unaware of or could not recall an assessment taking place it was unlikely to have been of value. The provision of feedback in this situation would be expected. There was therefore a significant group of trainees where there was inadequate assessment on a quantitative basis. It may be that the overall proportions using the assessment methods were lower than our figures suggest since we attributed the same amount of usage to users and non users. It is possible that non users were less likely to respond to the questionnaire.

The quality of the assessment carried out cannot be measured directly in this kind of survey but from the comments which appeared on many of the responses it appears that some trainers carried out the assessment with skill while others did not. In response to the question on video taping several

trainees commented on the fact that taping took place but nobody looked at the tape. There had clearly been a considerable increase in the use of videotape over the years. A retrospective survey (Kelly and Murray, 1991) has indicated that in 1979 only 5% of trainees in the region were using video or audio taping.

It was very encouraging that those who had experience of the assessment techniques rated them more useful than those who had not. This could perhaps be expected in the trainers who chose to use the techniques but it is significant that among the trainees only Video and OSCE reached the mid point of the scale among non users whereas among users only one method (Manchester Rating) remained below the mid point. This study indicated that if trainers and trainees can be encouraged to use assessment methods they will find them helpful.

It was interesting that while 30% of trainers claimed to use a standard assessment programme only 8% of trainees were aware of this. Since assessment of trainees should have a formative function it is essential that trainees should know if they are being assessed, how they are being assessed, and the results of the assessment. It seems unlikely that if an assessment

programme had been discussed with the trainees they would have no recollection of this within three months of completing the trainee year.

The question of a regional programme produced a fairly even split among respondents with just over half of the trainees in favour and just over half the trainers against. It might be thought that those trainers who already had an extensive assessment programme would be particularly opposed to a regional system but in fact there was no such correlation. Similarly those trainers who did very little assessment showed no particular enthusiasm for a regional scheme.

In its instructions to examiners in 1978 the council of the RCGP stated that "the examination must now be regarded as a method of assessing the satisfactory completion of vocational training". More than half of the trainees and nearly half of the trainers in the survey population did not feel that the MRCGP exam was a valid assessment. The most common reasons advanced for this view were the lack of a clinical component and a belief that the exam represented only one view of the characteristics of a General Practitioner.

## Conclusions

As a result of this survey it became clear that those trainees in Practices where a variety of assessment techniques are used appropriately believed that continuous assessment was of value. It is also clear that a minority of trainees were being denied such assessment. While the voluntary use of assessment had produced some benefits it appeared clear that universal assessment would only become a reality when it became a requirement for accreditation as a Trainer. In the light of this we proposed that the regional policy on assessment be formalised as follows:

Every trainee should have teaching on videotaped consultations at least three times in the year. An MCQ should be used at least twice a year. Confidence rating lists should be used at least twice a year. A training programme on the use of the OSCE was commenced with the objective of running an OSCE in each district within the region. It was also decided to examine more closely the pre-existing process of summative assessment whereby the trainer was solely responsible for certifying the trainees competence to practice independently. It was clear from this survey that a significant number of trainers had carried out insufficient assessment to enable them to come to a considered and evidence based judgement on this issue.

As a result of the findings in this study it was decided to introduce additional elements to the assessment process including a region wide formative assessment and several externally assessed elements in an attempt to produce a valid, reliable and credible summative assessment package.

**TABLE 1 - Recorded use of assessment methods**

Method	Number (%) of respondents reporting use of method	
	Trainees	Trainers
Videotaped Consultations	50(76%)	95(76%)
Trainer sitting in	33(50%)	86(68%)
Manchester ratings	33(51%)	77(62%)
Use of checklist	41(63%)	107(86%)
Use of written work	35(54%)	82(65%)
Use of OSCE	29(45%)	65(52%)
Use of hard data	17(27%)	60(48%)
Other methods	12(18%)	55(44%)

Note: not all trainees answered all of the questions so for some the total number of responses is less than 72.

Figure 1

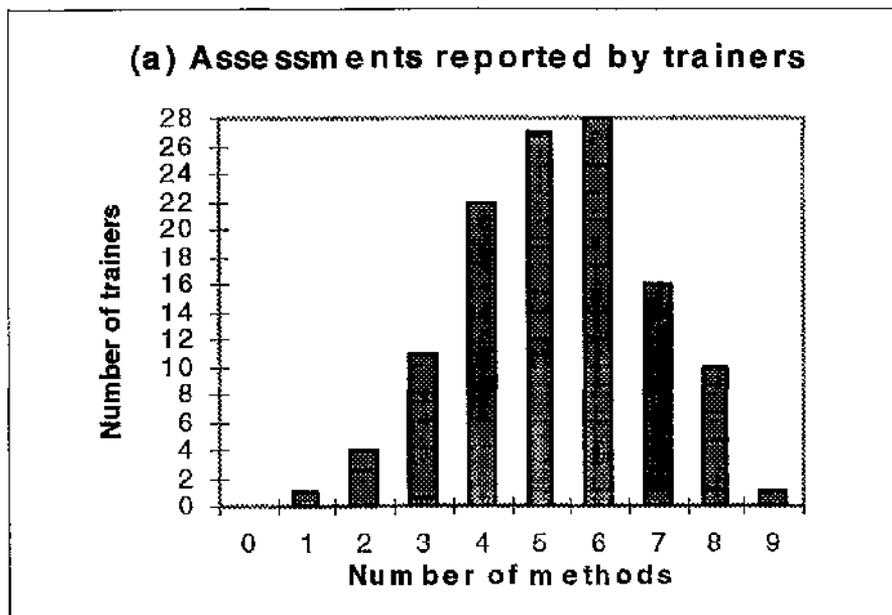


Figure 2

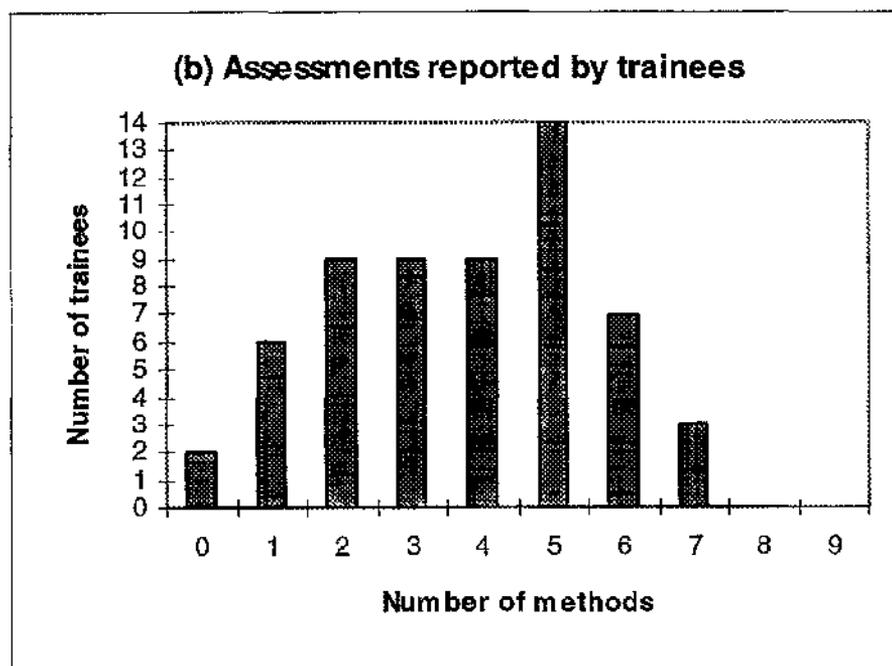


TABLE 2

Trainers' ratings of the usefulness of assessment methods					
Method	Mean (SD) rating by the trainers			Users versus non-users	
	All	Users	Non-users	$\chi^2$	
Videotaped consultations	4.0 (0.9)	4.2 (0.8)	3.1 (0.8)	37.5	P<0.001
Trainer sitting in	2.8 (1.3)	3.2 (1.1)	1.9 (0.7)	32.4	P<0.001
Manchester rating scales	3.1 (0.8)	3.3 (0.8)	2.7 (0.7)	18.6	P<0.001
Topic checklist	3.7 (1.0)	3.8 (0.8)	2.3 (0.9)	36.3	P<0.001
Written work	3.6 (0.9)	3.8 (0.8)	2.7 (0.8)	27.1	P<0.001
Objective structured	3.6 (0.9)	3.8 (0.8)	3.4 (0.9)	9.3	NS
Clinical examination					

SD = standard deviation. NS = not significant.

TABLE 3

Trainees' ratings of the usefulness of assessment methods.					
Method	Mean (SD) rating by the trainers			Users versus non-users	
	All	Users	Non-users	$\chi^2$	
Videotaped consultations	3.7 (1.0)	3.9 (0.8)	3.0 (1.1)	11.3	P<0.025
Trainer sitting in	2.8 (1.3)	3.3 (1.2)	2.4 (1.3)	16.0	P<0.01
Manchester rating scales	2.9 (0.9)	2.9 (1.1)	2.9 (0.7)	8.7	NS
Topic checklist	3.4 (0.8)	3.7 (0.8)	2.9 (0.3)	17.6	P<0.005
Written work	3.5 (0.9)	3.8 (0.9)	2.9 (0.7)	12.6	P<0.025
Objective structured	3.5 (0.9)	3.8 (0.8)	3.1 (0.6)	21.5	P<0.001
clinical examination					

SD = standard deviation. NS = not significant.

**TABLE 4**

Responses to questions on standard assessment programmes, regional programmes and the MRCGP examination		
	No. (%) of respondents <sup>a</sup>	
	Trainers	Trainees
With standard programme in practice	36 (30)	5 (9)
Approving of a regional programme	57 (48)	36 (57)
Regarding MRCGP exam as valid assessment	65 (53)	29 (44)

<sup>a</sup> Not all respondents replied to these questions.

## **CHAPTER 2**

### **Summative Assessment - The West of Scotland Pilot project**

#### **Introduction**

As a result of the work carried out as described in Chapter 1 it was decided to explore the possibility of developing a valid, reliable and feasible method for the summative assessment of general practice trainees.

Considerable work has been carried out on assessment of trainees in General Practice. The emphasis has been on assessment as an aid to teaching i.e. formative assessment. Mullholland and Tomblinson (1990) have produced a useful theoretical analysis of summative assessment. The General Medical Services Committee (GMSC), the JCPTGP and the Royal College of General Practitioners (RCGP) issued a joint statement in 1990 (Irvine) stating that the certificate of satisfactory completion of vocational training was in fact a certificate of competence, as opposed to a certificate of attendance. The following is a quotation from the Joint Committee on Postgraduate Training for General Practice (JCPTGP) document "Guidance on the Joint Committee's requirements for Regional and Scheme Accreditation:

“The Committee accepts the need for a national standard of entry into general practice and, therefore, the necessity to apply a system of assessment which is credible to both the public and the profession. Such a standard should reflect the attainment of the attributes of the general practitioner through the satisfactory completion of training and its achievement should be demonstrated by a competent system of assessment applied nationally.”

A survey of GPs carried out by the General Medical Services Committee (1992) 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 practice as a GP”, while 44% disagreed with this statement.

In view of these developments 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. This chapter describes the work of that group.

The pre-existing regional assessment programme

The West of Scotland Region had on average 155 trainers with around 150 trainees in post at any one time. A survey of existing local assessment procedures was carried out in 1989 (Campbell and Murray, 1990) as described in Chapter 1. This demonstrated extensive but patchy use of assessment methods throughout the region. As a result of this survey the West of Scotland introduced a regional formative assessment package. The following example is for a typical trainee starting the general practice year on the 1st of August:

Annual Assessment programme:

<b>August</b>	Confidence check list, Multiple choice paper (MCQ)
<b>September</b>	Formative video assessment
<b>October</b>	Trainee interview with Associate Adviser
<b>November</b>	Manchester Rating, Objective Structured Clinical Exam (OSCE) (Harden 1979)
<b>January</b>	Formative Video Assessment
<b>February</b>	Check list, MCQ
<b>March</b>	Manchester rating, OSCE
<b>April</b>	Formative video assessment

This programme became mandatory from 1990 for all training practices in the Region. The package was designed as a formative procedure and was felt to have serious shortcomings as a potential summative assessment process.

### Standard setting

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 agreed by our working group which we considered to be reasonably comprehensive.

### Proposed attributes of a summative assessment programme

- 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
- A 100% pass rate should be possible
- The procedure must be feasible

Our group took the view that all of the above were essential prerequisites for a summative assessment process for the following reasons. The trainer is the person who has most 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 the trainer to form an objective and unbiased view of trainee performance. Additionally it could be felt that failure of a trainee to reach a satisfactory level of competence casts doubt on the trainers ability to appoint an appropriate trainee or provide adequate teaching. The possible conflict of interest is such that an external contribution to the assessment process will be necessary for the credibility of the system. For any assessment method to attain face validity it must measure an area which is relevant to the eventual professional activities of the candidate. Few would dispute the concept that the main role of the General Practitioner is to provide continuing health care for individuals and families. The ability of the doctor to carry out consultations successfully is therefore a major determinant of the doctors overall competence.

It has been shown for example (Millar, 1991) that GP trainees with good interviewing skills are more likely to offer relevant advice and treatment to

patients with psychiatric disorders. 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 (Rabinowitz, 1986). In addition MCQs may affect the students' approach to learning producing a superficial rote learning approach. Individuals have a varying response to the situation of an end-point exam. There is considerable anecdotal evidence that some candidates do not perform well under exam conditions. Some element of continuous assessment would therefore appear to be desirable.

It would be inappropriate in an assessment of this nature to have a built in failure rate as occurs in many postgraduate exams. Clearly it should be possible for all candidates to pass. This entails the use of criterion referencing rather than peer referencing wherever 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. We have chosen to take the latter course. There is some evidence that such global scales are at least as

reliable as more complex marking systems. (Roberts (1990) has shown that although raters can be very reliable using detailed marking systems interstation reliability is very low. This is due to different skills being assessed at different stations rather than any inherent variability in performance.

### The Legal position

The legal position at the time was that the trainer and relevant hospital consultants have the statutory responsibility to sign the Certificates of Completion. The West of Scotland working group took the view that in the short term there need be no change in this system. It was felt that the summative assessment results should guide the trainer in making this decision. Clearly this has an air of coercion about it but it is the duty of Regions to satisfy the JCPTGP 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.

### Summative assessment options

The working group identified three possible assessment systems which might be used in summative assessment:

- The status quo
- The Royal College of General Practitioners membership exam
- A specifically devised system

The status quo had some attractions. In the West of Scotland region the granting of a certificate is based on the trainers judgement informed by the current formative assessment schedule. Clearly the trainer is well placed to identify the strengths and weakness of the trainee. A further attraction is that to use the current formative system would incur no additional time or resources. However, it was felt that this approach was fundamentally flawed in that it would lack objectivity as well as credibility with the public and regulatory authorities because of the absence of any contribution from outside the practice.

We further felt that 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 exam has positive aspects. The examination has been developed over a period of years and its reliability has been extensively studied. However, the working group took the view that the RCGP exam as currently constituted had two significant drawbacks. Firstly there was no clinical component. This possible drawback had been noted as long ago as 1979 by Hannay. Secondly the RCGP exam is peer referenced in that the pass rate is effectively fixed at around 75% and does not vary from year to year. Additional problems with the College exam were the lack of trainer input and the obvious connection with the RCGP a body to which less than half of GP Principals belong. 32% of GP Principals in the region are college members. The use of the RCGP examination would also entail using an instrument designed for one purpose (assessment of suitability for college membership) and using it for an entirely different purpose (identification of acceptable competence).

It was decided therefore to attempt to devise a region based system which would satisfy all of our criteria.

### The summative assessment methods

The working group decided to use four components in the assessment process:

- A multiple true/false paper (MCQ)
- Trainers overall judgement
- A completed audit
- Assessment of videotaped consultations

Trainees identified by this process as being of doubtful competence would be referred for further assessment as discussed later. 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 MCQ is a reliable and feasible method of identifying factual knowledge. One Region in England requires a minimum acceptable performance in an MCQ before a potential trainee will be accepted into a practice (Walker, 1987). The trainer is uniquely placed to observe the trainee over the course of the year particularly in the areas of attitudes and behaviour. Performance review has long been recognised as a necessary component of the practice of medicine. A completed audit will

demonstrate that the trainee has absorbed the principles and carried out the practice of performance review.

The major part of general practice takes place in the consultation with an individual patient. Therefore for any system of assessment to be credible it must address this area. We identified three possible approaches to assessing performance in the consultation:

- Direct observation of consultations
- The use of actors in a simulated surgery
- Video-recordings of real consultations

Direct observation by an assessor has several attractions. The assessor can attempt to confirm the trainees findings by taking additional history and examining the patient, discussing the trainees actions with the trainee and obtaining immediate feedback from the patient. However, the group took the view that there would be considerable problems with logistics in that an assessor would have to spend a session with each of 10 trainees. In addition there are obvious potential problems of disruption of the consultation by the observer. Finally, there would be no possibility of assessing the observers

performance unless the consultations were recorded in addition to being observed.

Much work has been done on the use of the Simulated Surgery and its earlier version the Objective Structured Clinical Examination (OSCE). The OSCE has been used on a region wide basis both in the West of Scotland and elsewhere. It does provide the opportunity to present the trainees with a group of standardised presentations. However, to carry out a simulated surgery for 150 trainees with 6 patients per trainee would require 300 assessors (assuming double marking) and 60 patients with each patient performing 15 times. The numbers of assessors and actors could be reduced by spreading the process over several days but the overall time element would remain the same and there would be additional problems of contamination.

Continuing care is a major concern in general practice and here the simulated surgery with its procession of new patients can be of little value.

Much of the world-wide work on consulting and clinical assessment has been based on simulated or standardised patients as first described by Harden in 1979. The main advantage of using this technique is that each candidate is

presented with the same set of problems to deal with. This is particularly important if it is wished to rank order candidates since the candidates are competing over the same course. Using this system impressive reliability figures can be produced provided there are sufficient cases used but this requires a test length of 6 hours and the reliability figures disguise large differences in station to station performance (Newble and Swanson, 1988).

However, in most cases simulations have been used to assess specific skills rather than overall competence (Ferrell, 1995; Williams, 1987) and the validity of simulations in complex performance has been questioned. Cox K (1990) has suggested that the use of the Objective Structured Clinical Examination is being extended inappropriately into final clinical and postgraduate examinations. He states that rigid examination structures are inappropriate for clinical tasks requiring eclectic, responsive skills controlled by clinical judgement. As an alternative he advocates performance assessment based on day to day clinical tasks. Although the clinical process can be divided into parts which are sufficiently simplified and specified to enable accurate measurement, he proposes that it does not follow that measuring the separate parts is equivalent to measuring the whole integrated performance. It is pointed out that clinicians do not work systematically

through a sequence of routines, but search for clues to fit explanations that spring to mind as the picture develops (Elstein, 1978). Cox claims that what distinguishes experts from novices is a broader repertoire of approaches. Scoring systems based on a fixed set of components penalise the shrewdness and efficiency of those who reach a diagnosis with the least number of tests in the shortest time. The main concern using real patients is that the cases will be inappropriate for making judgements with for example too many low challenge consultations or too many of the same type of case. However, work has been reported using genuine consultations (Hays, 1990; Cox J, 1993). Real consulting sessions have advantages over simulations in that the trainee can record the consulting session at a time of his/her choosing and the stress of an exam situation is removed since the trainee can try again if the first session is unsatisfactory.

The use of videotaped consultations has other potential advantages. It enables observation of real consultations in a relatively unobtrusive way. Tapes can be assessed by a number of assessors thereby measuring the reliability of rating and also calibrating the assessors. A further advantage is that by allowing the trainee to record several sessions and then submit a session with which he/she was happy we would eliminate the problems of

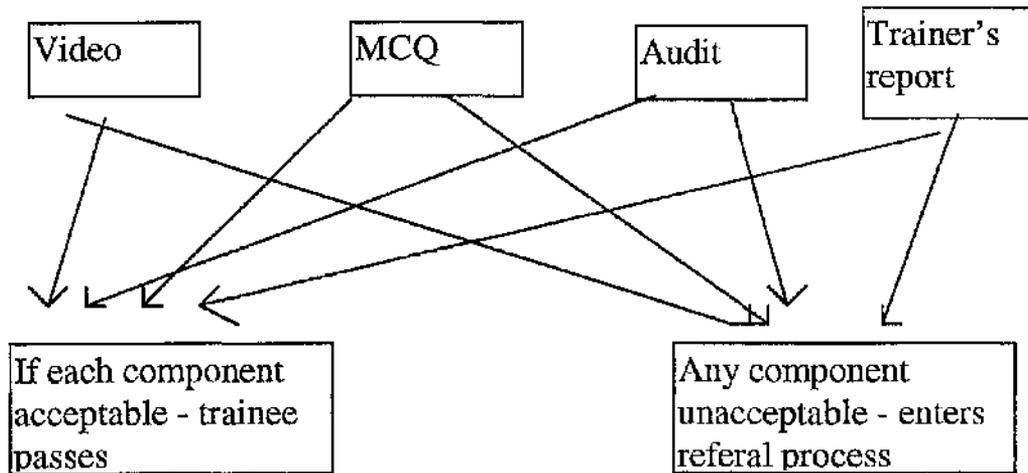
exam nerves and 'performance anxiety'. Video also appeared to the working group to be the most logistically feasible option. Potential difficulties with regard to technical factors, confidentiality, reliability of assessment, case mix, and level of difficulty of consultations were identified and will be addressed in the following pages. Despite these potential difficulties it was agreed that it was necessary to attempt to measure consulting competencies as described by Maguire (1989).

The group decided that the number of assessors involved should be kept low to reduce variability but should contain enough people to keep the workload to manageable proportions. It was agreed to invite applications from Associate Advisers, members of training practices, and others with suitable assessment experience. In the event 25 assessors were appointed. All applicants were accepted. The majority were trainers, 4 were course organisers, and 7 were current RCGP examiners.

### The assessment path

A trainee who was rated as satisfactory in all four components would automatically receive a certificate. A trainee who was rated doubtful in any one area would enter a referral process. This is shown in figure 3.

Figure 3



The referral process for each component will be discussed in detail later. The working group did not take a view on the number of trainees likely to fail to obtain a certificate under the new scheme but it was felt that it would be unlikely to exceed 5%.

### Videotape analysis

It was agreed that the initial effort would be devoted to the videotaped consultation assessment since the use of MCQs and Audit appeared to be more straightforward and the trainers judgement component was already in use. In Chapter 4 I will look at the audit and MCQ components in more detail

### Production of videotapes

155 Trainees were invited to produce 4 hour videotapes and accompanying log books. Within 6 weeks 80 videotapes and logs had been received at the Regional office. The first batch of tapes were sent to the assessors for practice purposes prior to the assessors workshop. By the time of the workshop all assessors had the opportunity to view at least two tapes. The assessors were asked to assess as many consultations as they felt they needed to come to a firm judgement but in any event a minimum of six consultations. Much of the workshop was devoted to viewing and analysing trainee tapes although the other 3 aspects of the assessment package were also considered. The main conclusions of the workshop were as follows:

### Production of videotapes and tape length

We initially asked trainees to produce 4 hour tapes but in none of the tapes viewed was this length felt to be necessary. A two hour tape containing at least two paediatric consultations and two chronic problems was felt to be adequate but further a further prospective analysis was planned. This will be discussed in Chapter 5.

### The Trainee log

An initial version of the trainee log worked reasonably well but it was agreed that an improvement would be to give the trainee the opportunity to grade the degree of difficulty of the consultation (see Appendix B) .

### The assessment form

Assessors were originally 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 et al. Health promotion was included in the first version of the form but this was found to be almost impossible to rate in many consultations. Health promotion is an important part of general practice and should be part of the summative assessment process. It was felt that this would be better assessed as part of the trainers report, with factual knowledge about health promotion being dealt with in the MCQ. The workshop also decided to add item 'O' to the assessment form. We felt that this would enable the assessor to register a serious error and then concentrate on the other aspects of the consultation.

The version of the form which emerged as the most valid and reliable in the opinion of the group was as follows:

O: Was there any obvious diagnostic or management error?

A: How clearly did the doctor discover the reasons for the patients attendance?

B: How clearly did the doctor define the clinical problem?

C: How well did the doctor tailor the explanation to the needs of the patient?

D: How well did the doctor manage the clinical problem?

E: How effectively did the doctor use resources?

F: How effectively did the doctor relate to the patient?

The use of these scales is discussed in Chapter 3 and the final version of the instrument which eventually came into use as the UK consulting assessment instrument is shown in appendix C.

#### Assessment reliability and assessor training

Assessors' opinions appeared to be reasonably consistent during the rating sessions in that there was a large degree of agreement on which consultations were satisfactory. However, at this stage no formal reliability work was

carried out. Reliability data will be presented in Chapter 3. The initial assessor training consisted of an explanation of the purpose behind summative assessment, an introduction to the marking instrument and trial marking of several consultations. It became clear that there were no black or white situations. The main issue for the assessors was whether a trainee performed adequately in the various parameters. There was an increase in convergence as more consultations were viewed but verdicts on individual consultations were rarely unanimous. In terms of overall decisions regarding trainee competence the assessors agreed on 90% who were felt to be competent and on 1-2% who were not competent. The remaining 7-8% produced differing views among the assessors. Further detailed work on reliability was organised as a result of this pilot. It was also decided that there should be regular meetings of assessors for calibration purposes in order to produce as much consensus as possible. The decision was made to keep assessor numbers as low as possible compatible with workload in order to minimise inter-observer variability and to give each observer sufficient tapes to facilitate self calibration. If we have 25 assessors looking at a total of 250 (125\*2) tapes each assessor would have to review approximately 10 tapes in one month. This would be expected to involve 10 hours of work. Although

this is a considerable workload it is important that each assessor sees a reasonable number of tapes for calibration purposes.

### Problems encountered

#### a) *Technical problems*

Some trainees submitted tapes on small VHS cassettes, others used super VHS. In total these only added up to 2-3% of tapes but required a disproportionate amount of time to transfer to standard VHS. It was clear that most Practices when buying a new camera were selecting a small format machine for reasons of cost and convenience. Picture and sound quality was reasonable overall but it was clear that some standard advice in these areas should be produced in any final instruction sheet. In particular the use of camera clocks and desk top microphones made the tapes much more assessor friendly.

#### b) *Objections in principle.*

The department received 12 letters from trainers and two letters from groups of trainees. The main objections advanced were as follows:

### ***The timing of assessment***

Some trainers took the view that incompetent individuals should be identified long before the trainee year. The working group agreed that this should be the case but felt that there is there was little evidence that this was actually happening.

### ***Concerns about confidentiality***

Some doctors were concerned about videotapes leaving the practice at all. Others were concerned that the tapes would be viewed by non medically qualified personnel. These problems had been anticipated and all tapes were sent by recorded delivery or by hand. The Educationalists working with the group were bound by codes of confidentiality in the same way as doctors.

### ***Concerns regarding consent***

The postgraduate office 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 was submitted to the Defence Societies and the West of Scotland General Practice ethical committee whose responses were favourable. Largely as a result of work on video for summative assessment the GMC produced guidelines for

the recording of videotapes. The latest version of our consent form complies with these guidelines and is reproduced as Appendix E. A guide to trainers on informed consent and an explanation for patients was produced for the full scale version of the programme.

### ***Disruption of the consultation***

Some disquiet was expressed that the presence of the camera would disrupt the consultation and prevent the patient discussing problems. A study by Martin and Martin (1984) had shown that 78% of patients who agreed to video recording forgot about the camera during the consultation while 16% felt that the camera made them less willing to talk about embarrassing topics. 92% of patients felt that video recording was a good way for doctors to look at what they are really doing to patients. We were conscious that adverse effect on the consultation from the patient's point of view would be a considerable drawback to the use of video and set up a study to look at this area. This study is reported in Chapter 7.

### ***Lack of consultation***

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 unrest, but the working group decided that full consultation was not possible in view of the speed with which events were happening.

### Preliminary results of videotape assessment

Does viewing consultation tapes help us to decide competence?

Our 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 added an additional dimension to their ability to identify trainees' consulting abilities. We did, however, encounter some difficulties which did indicate that video could not give us all the answers. We had originally included health promotion as one of our parameters but found it impossible to decide in many cases whether or not appropriate health promotion had been offered. The reason for our difficulty was that the amount of health promotion offered would depend not only on the nature of the presenting problem but also on the amount of data already in the patients records. For example questions about cervical smears, smoking, blood pressure etc. would not be relevant if the patient was known to be non

smoking, normotensive with a recent normal smear. This kind of information was obviously unavailable to us.

The other attributes were usually consistently assessed but we did have difficulty on occasion where we felt the candidate had made an incorrect diagnosis but to be certain we would have had to examine the patient, which was clearly not possible.

On a positive note we identified unexpected areas about which the videotape gave us information. The degree of organisation of the individual trainee and the practice was often apparent. A remarkable number of trainees did not refer to the patients notes prior to the consultation although they may have gone through all of the records prior to the start of the session. A significant number did not refer to the notes during the consultation. This produced results which were sometimes unfortunate. Trainees were heard to ask patients their identity, and to ask questions which were already answered in the records. More worryingly some patients had to point out to doctors information which the doctor required to use, such as a recent abnormal smear.

### Future Plans

A formal analysis was planned to assess the reliability of the videotape assessment using two assessors per tape. An attempt was planned to look at the number of consultations to be viewed by asking our assessors to rate the degree of impact of succeeding consultations. The group also hoped to identify which of the parameters in the rating scale were marked most or least consistently. In addition we intended to look at possible correlations between performance on video with knowledge base using MCQs and with performance in the RCGP examination.

The Joint Committee Working Party had called for feasibility studies in summative assessment. In the event that summative assessment became mandatory the group hoped to have the package available in a validated form in sufficient time to ensure that any developments are instituted by the profession rather than imposed by the regulatory authorities.

### Conclusion

Our working group of experienced General Practitioners came to the provisional conclusion that the use of videotaped consultations may well be a

valid and reliable method of assessing the competence of General Practice Trainees as part of a balanced regional summative assessment package.

### Implementation

The date for full regional implementation was set for August 1994. It was felt that the package could be implemented in stages as the assessment tools were developed. The trainers assessment was already in place and a full scale pilot of all four components was organised to be run using those trainees completing training in July 1993.

## **CHAPTER 3**

### **The use of videotaped consultations in Summative Assessment of Trainees in General Practice**

#### **Introduction**

In an attempt to formulate a valid, reliable and externally credible summative assessment programme the West of Scotland Committee for Postgraduate Medical Education developed a system based on four components:

- The trainers overall judgement
- Assessment of videotaped consultations
- A multiple choice paper
- An Audit project

In Chapter 2 I have described the rationale behind the adoption of these components.

This framework has been adopted by the summative assessment working group of the JCPTGP. Failure by a trainee to satisfy the assessors in any one

of these components initiates a referral process in which the evidence is reviewed by additional assessors from outside the region. In Chapter 2 we have described the system as a whole and the development of the marking schedule and rating scales to be used.

The use of videotaped consultations as an educational resource is well established (Pendleton, 1984) and we have shown (Campbell and Murray, 1990) that trainees find this use of value. Several instruments have been developed to rate consultations (Hays, 1990; Fraser, 1992 and 1994; Cox, 1993). However, none of these scales has been used in summative assessment of vocational trainees. We investigated the possibility of using the scales developed by Cox and by Fraser (The Leicester assessment package). However, we identified major difficulties in the potential use of either of these rating scales in the summative assessment of vocational trainees. The scale developed by Cox is designed for educational assessment but had no cut off point to denote minimal competence and was cumbersome to use. There were difficulties also in identifying minimal competence using the Leicester assessment package (Campbell and Murray, 1994). Professor Fraser's group stated that candidates scoring less than 50% should be regarded as of unacceptable competence. We could not find a pass/fail mark

for summative assessment in the LAP. In the LAP it is stated that a score of below 40% demonstrates that the doctor is not safe to practice independently - below which figure would trainees be refused a certificate? In addition the LAP scoring system is such that it is possible to be totally inadequate in one component yet still score over 50%. Similarly there is no system for 'black-balling' a candidate who makes a single gross error.

In the Leicester study itself none of the doctors had an overall mean score below 50% i.e. none were of unacceptable competence. The authors were therefore claiming that their system could identify a group of doctors none of whom had actually been tested by the system. Although, accepting Fraser's definition of minimal competence, two out of six assessors believed that one doctor fell below this level while the other four disagreed. We took the view that such disagreement must raise doubts about the utility of the method in summative assessment. Although it is not explicit in the text of the LAP reliability study it appeared that this doctor had no experience at all of general practice but still managed to produce 'an acceptable performance in general practice consultations'. If this result is repeatable it must raise questions either about the validity of the assessment method or the need for vocational training.

The principal objective of this study was to determine if videotaped consultations could be used to help to identify the small number of trainees who may not yet have reached acceptable levels of competence.

The specific objectives were:

1. To assess the practicalities and acceptability of tape production
2. To identify whether routine consultations were sufficiently challenging to assist assessors to differentiate competent from non competent trainees.
3. To test out the marking schedule.
4. To measure inter-observer reliability.
5. To identify the number of consultations and assessors needed to assess each trainee reliably.

### **Method**

A letter was sent to the one hundred and fifty practice based Trainees in the West of Scotland inviting them each to submit a 4 hour videotape of routine consultations. 4 hours was chosen since it was felt to be probable that the

time needed would be less than this but how much less was unknown. The letter contained advice on the techniques of obtaining videotapes of suitable quality and also covered such areas as obtaining informed consent from patients.

The trainees were also provided with a simple log for the recording of details of each consultation. The function of this log was partly to enable assessors to find their way through the tape but also included a section in which the trainees were given the opportunity to discuss how the consultation was handled. A total of 80 tapes were received within six weeks of the invitation.

### *Methods*

Ten tapes were selected for the study. In order to provide a test of the reliability of the process at the level of questionable competence, one tape was specifically selected from a trainee who was causing some concern to the trainer. The remainder of the tapes were chosen randomly. With all of the tapes it was recognised in that many of the assessors were involved in training and might have other information about trainees which could affect their decision making. However, the West of Scotland is a large region and

in practice assessors would not be asked to view tapes of trainees from the same locality.

The 25 assessors used were the same group who had taken part in the original pilot project. The majority (14) were trainers, four were associate advisers (the equivalent of course organisers in England) and seven were examiners for the RCGP. Each of 25 assessors was sent the ten videotapes with accompanying log books in which the trainees were asked to record factual details about the consultation plus any comments felt to be appropriate. Assessors also received assessment forms for each trainee. The assessors were allocated four weeks to complete the assessment. Twenty assessors returned the completed assessments in time and the results are based on this group. A further four assessors eventually completed the task. One assessor dropped out at this stage for unspecified reasons. Assessors also received assessment forms for each trainee. Assessors could record specific details of the consultation under strengths and weakness; no score was attached to these evaluations but the assessor could use the comments as an aide-memoire when reaching a final decision.

Assessors were asked to rate the trainee's performance in each consultation in seven areas:

- 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 well did the doctor use resources of time, investigations and manpower?
- How effectively did the doctor relate to the patient?

All these attributes apart from the first were then scored on a six point scale as follows:

1 = Refer

2 = Probably refer

3 = Bare pass

4 = Competent

5 = Good

6 = Excellent

The word 'refer' was used rather than 'fail' since an unsatisfactory performance in the video component would lead to a further assessment rather than to a refusal of the certificate of satisfactory completion. Assessors were instructed to view as many consultations as were necessary to reach a final decision but in any event a minimum of six. In addition, the assessors were asked to rate the degree of difficulty of each consultation, to form a judgement on each consultation, to record an overall cumulative opinion of the trainees competence after each consultation and to record the degree of impact each consultation had on their overall opinion. When assessors reached a firm conclusion they were asked to record a final judgement which had to be either refer or pass. It was recognised that an assessor, having noticed some aspect of performance, might wish to examine a consultation of a particular type to help clarify judgement. For this reason assessors were not restricted to watching the same series of consultations.

All statistical calculations were carried out using the statistical package *SPSS-PL VERSION 4.01*. In order to determine the relationships among the six attributes on the rating scale (apart from the error attribute) a principal components analysis was carried out. All consultations were used for this

analysis giving a total of 1176 events. The principal components analysis determines how closely scoring in any one item is related to scores in the other items. In order to determine inter-rater reliability, the score given by each assessor to the first consultation of each trainee was used to produce a rank order of consultations by total score for each assessor. Correlations between assessors in terms of rank ordering were then assessed.

## **Results**

### *Assessment scales*

Eight assessors recorded that they believed an error in diagnosis or treatment to have occurred in the case of one particular trainee. Errors were also recorded once for each of two other trainees. Other than this no specific errors were recorded.

The interrelationships among the 6 sub-scales on which each consultation was rated were first examined. The Correlations among these are shown in Table 5. A principal components analysis was performed which confirmed a single underlying factor (explaining 76.3% of the variance overall). This would indicate that each of the items is in fact measuring a different aspect of the same overall behaviour pattern. Composite 'factor scores' for each

assessment of a consultation were thus calculated, and used subsequently. These may be seen as overall judgements about the consultation. These overall scores were subsequently used in comparing assessors behaviour overall. The overall mean score given by all assessors was set at 0. Table 6 shows the number of consultations where complete scores were recorded, and the overall mean score given to all consultations assessed. A positive score indicates high marking relative to the other assessors and a negative score indicates overall low marking. For example assessor 18 had the highest overall mean score but referred one trainee.

### *Consultations*

Not all assessors elected to view the first consultation on each videotape consequently the results are based on 18 assessors. Correlations between assessors in terms of rank ordering were poor with the correlation for individual assessors with the mean ranking ranging from 0.2 to 0.5. This suggests that using the scoring system arithmetically to reach overall judgements would be unreliable although correlations would improve with increasing numbers of consultations.

### *Trainees*

Table 6 shows the number of trainees referred by each assessor and Table 7 shows the number of times individual trainees were referred. Five assessors did not refer any trainee and no assessor referred more than two trainees. On the other hand, 15 assessors referred one particular trainee. One trainee was referred by two assessors and two trainees were each referred by a single assessor. Using a simple odds ratio calculation, with one assessor the first trainee would have 15 chances out of 20 of being identified, and with two assessors the combined probability of identification becomes 95 out of 100. Thus, using two assessors for each videotape the first trainee would have a 95% probability of being referred; no other trainee would have had more than a 20% chance of being referred after a minimum of six consultations.

### *Challenge*

The number of consultations rated for challenge is not identical to the number of consultations assessed as some assessments had data missing and were excluded from the calculation used to calculate scores (Table 8). Assessors varied in how challenging they perceived consultations to be for the trainees. For example, assessor 1 rated 76.6% of consultations as being of low challenge while assessor 7 rated 9.4% as low challenge.

### *Cumulative rankings*

In order to determine how many consultations needed to be assessed, the occasions where assessors made an important change in their overall opinion, i.e. from pass to refer or vice versa, were analysed. After four consultations no assessor changed an overall judgement in such away.

### **Discussion**

The use of real consultations as part of a process to assess the fitness of trainees to receive a certificate of satisfactory completion has obvious face validity. There is corresponding evidence that paper based assessments using multiple choice or modified essay papers are not good predictors of actual performance (Rabinowitz, 1987). An alternative to real consultations is the use of simulated patients (Tamblyn, 1991; Vu, 1992; Harden, 1979; Norman, 1985; Rethans, 1987; Colliver, 1991) but problems of patient consistency have been noted using this method (Tamblyn, 1991; Vu, 1992).

The carrying out of the videotape assessment by assessors from outside the practice has considerable potential benefits with regard to objectivity and external credibility. However, in order for this form of assessment to be

worthwhile it must produce results of adequate reliability while taking up only a reasonable amount of trainee and assessor time.

### *Tape Production*

90% of tapes produced were technically usable. Trainees appeared to have no difficulty obtaining appropriate patient consent.

Disquiet was expressed by some trainers, both to the authors and to medical journals about consent, confidentiality and the effects of video recording on doctor and patient behaviour (Baird and Gillies, 1993). A survey has shown that a number of patients (none of whom had ever been videotaped) felt unhappy about their possible reactions to such a request (Bain and Mackay, 1993). However, other work has shown that most patients are happy to give consent to video recording (Martin and Martin, 1984) and that Doctors are unaffected by the presence of the camera (Pringle et al, 1984; Pringle and Stewart-Evans, 1990). Clearly it is important that it should be as easy as possible for a patient to withhold consent or to have the tape erased after recording. At the time of this study we used the most authoritative guidelines available (Southgate, 1993). The General Medical Council have now produced definitive guidelines (1995).

Cameras were available to all Trainers groups and many practices had their own cameras. 90% of tapes were technically usable. No trainee reported any difficulty gaining access to a camera. The sound quality was usually adequate but was much enhanced by the use of a desk top microphone.

In order to complete the log it was necessary for the trainee to view the tape thus taking up a maximum of 4 additional hours. In view of the large number of trainees who produced tapes the process was clearly practicable for the trainees.

### *Workload*

The 25 assessors used were all volunteers who were principals in general practice. The 24 assessors who completed their marking tasks found the workload acceptable. Only one assessor dropped out. The assessors spent an average of 10 hours carrying out the process. This would be same amount of time required to assess all trainees in the region using 2 markers per trainee (140 trainees with each tape viewed by two assessors out of the 24 producing a total of 10 tapes per assessor). All assessors who completed the task expressed a willingness to continue in post.

### *Interpretation of Results*

Assessors were deliberately allowed to select for themselves which consultations to view. This has advantages in that assessors can attempt to seek out particular areas of competence to examine. However, this approach does produce difficulties in the analysis since it introduces an additional variable in that not all assessors looked at the same consultations.

It is clear from the correlations between the different components of our rating scales that although the six point scale may have been helpful in directing the assessors to look at these areas the assessors' judgement of trainee performance in a given consultation was largely consistent across the different parameters. This finding is understandable since minimal competence is less likely to be case specific than higher order skills. For example, it seems reasonable that a trainee who attempts to take an adequate history will also attempt to explain the problem appropriately. The six component scale was therefore not helpful in increasing reliability. The use of the six point scale was therefore modified for the reasons discussed above and for future use has been replaced by a simple 3 parameter scale to help assessors focus on important areas:

- A. Listening - Did the trainee identify adequately the patient's problems?
- B. Action - Did the trainee investigate/manage the patient's problems appropriately?
- C. Understanding - Was the trainee aware of the strengths and weaknesses of the consultation?

In the detailed assessment instrument the assessors were provided with expanded versions of these criteria to enable them to judge if the registrar had achieved the criteria successfully. The word pictures used were gradually modified in the light of extensive usage. The final version as used in the national summative assessment programme is now as shown in appendix C. Negotiation was specifically mentioned in the this revised version. It should be borne in mind however that the word pictures used are to guide rather than direct the assessors. The assessors target remains to identify any GP trainees who may be performing below acceptable levels in consulting.

Item C would be completed in the light of the trainees comments in the log book. The revised scales were then tested out with the succeeding cohorts of GP registrars.

As stated earlier the inter observer reliability for individual consultations was not impressive but this observation is based on small numbers of events compared with the total events of 1176. It has been shown elsewhere that large numbers of consultations would be required to produce reliable results using a scoring system (Cox and Mulholland, 1993). As Table 6 demonstrates examiners varied considerably in the scores given to each candidate and in their range of scoring. However, there was much more agreement among assessors with regard to the overall decision to pass or refer. Using two markers per tape trainee number 1 would have had a 95% probability of being referred, no other trainee would have had more than a 20% chance of being referred. Following discussion among the assessors this trainee was considered to be just below an acceptable level of competence. This would indicate that the 95% probability of being referred would apply to all trainees who fell below the level of acceptability.

It is clear that our assessors were unable to come to a reliable decision concerning the competence of a trainee on the strength of viewing any single consultation. There are obvious reasons why this should be true. It is very unlikely that any single consultation will test out the range of competencies

needed in general practice, indeed a low challenge consultation may require very little in the way of competence. The reality that different competencies are required in different consultations has implications for any rating system which involves giving scores for each consultation and then producing an aggregate mark by some form of arithmetical manipulation (Fraser, 1992; Cox, 1993; Hays, 1990). Since our rating system involved recording interim judgements after each consultation but then producing a final decision based on the cumulative impression this difficulty has been avoided.

Of course the fact that our assessors were reasonably consistent does not mean that their decisions were correct. Studies of outcome validity would require the long term follow up of large numbers of trainees who had been deemed competent by different assessment methods. The undesirable endpoints would might include proven breaches of terms of service or findings of professional misconduct by the General Medical Council, events which are fortunately still relatively rare. However, the fact that by defining competence in terms of what the normal general practitioner perceives to be competence we used the same definition of competence as do the courts and the GMC, leads us to hope that this process will have positive predictive ability.

It is of interest that the assessors showed considerable variation in the degree of challenge they ascribed to the consultations viewed. A possible explanation of this is that the trainees were able to accept or reject potentially challenging situations. Some of our assessors may have been evaluating the potential challenge of the consultation while others may have based the judgement on the explicit challenge contained in the consultation which actually took place. 52% of consultations overall were rated to be of at least moderate challenge. This would indicate that routine consultations are of sufficient challenge to enable assessment of performance to take place.

When the number of consultations required to be viewed before a confident judgement could be made was studied, there was no occasion where an examiner changed from pass to refer or vice-versa in the seventh or subsequent consultation. As can be seen the majority of assessors decided that after 6 consultations they had seen enough to make a firm judgement. It is worth noting that assessors tended to view a larger number of consultations in situations where the eventual decision was to refer, although in no case did this extra viewing change the result.

The question arises as to what would happen if trainees were to attempt to produce a videotape consisting of 'good' consultations. Clearly if trainees are to be refused a certificate of satisfactory completion as a result of the process there would be an incentive to attempt to edit videotapes in this way. In a later chapter I will report on the first three years experience of the system in use and will discuss this possibility. However, it is important to remember that competence as opposed to performance is being assessed. True performance, as discussed in the introduction could only be assessed if the trainees were unaware they were being assessed; there is a well recognised difference between what doctors can do as opposed to what they routinely do.

Although the assessors showed limited agreement on the individual components of rating scales and on their ratings of individual consultations, they nevertheless showed an acceptable level of agreement on the ultimate issue of whether or not the trainees were competent, and their decisions became stable after observing a maximum of four consultations.

A continuing monitoring programme to identify those examiners whose results do not correlate well with their peers is needed to improve reliability as is further training and calibration for our assessors. The use of two

assessors per tape produced adequate reliability with a feasible workload. The region as a result of this pilot proceeded to carry out a full scale pilot examination involving all trainees finishing training using the revised assessment schedules with two assessors per trainee.

The main alternative to the use of videotaped consultations in summative assessment of consulting competence would appear to be the objective structured clinical examination or its variant the simulated surgery (Harden, 1979; Tamblyn, 1991; Vu, 1992; Norman 1985). Evidence concerning the use of these methods in summative assessment has not so far been produced.

**TABLE 5**

## Correlations among sub-scales

	A	B	C	D	E	F
A Score	1.00					
B Score	0.74	1.00				
C Score	0.69	0.75	1.00			
D Score	0.70	0.77	0.79	1.00		
E Score	0.66	0.68	0.69	0.76	1.00	
F Score	0.69	0.69	0.74	0.70	0.67	1.00

Number of cases : 1176, 1-tailed signif:p<.001 for all correlations

**TABLE 6**

	Number referred	Overall score	SD	Cases
Assessor 1	1	-.17	1.03	46
Assessor 2	0	.76	.64	60
Assessor 3	0	.49	.82	60
Assessor 4	1	.06	.93	66
Assessor 5	2	-.08	1.28	53
Assessor 6	1	.37	1.22	62
Assessor 7	0	-.82	.77	62
Assessor 8	1	.14	.88	58
Assessor 9	1	.18	.57	56
Assessor 10	0	.18	.84	51
Assessor 11	0	.36	.52	59
Assessor 12	2	-.58	.99	58
Assessor 13	1	-.32	.87	61
Assessor 14	1	-.30	.73	74
Assessor 15	2	.09	.89	61
Assessor 16	1	-.21	.97	60
Assessor 17	1	-.09	.84	48
Assessor 18	1	.78	1.36	60
Assessor 19	1	.16	.88	67
Assessor 20	2	-.51	.76	54

**TABLE 7****Overall decisions by assessors**

Trainee	Number of refer decisions
1	15
2	0
3	1
4	1
5	0
6	0
7	0
8	0
9	0
10	2

**TABLE 8****Challenge**

Assessor	Low	Medium	High
1	36	8	3
2	9	49	8
3	10	39	11
4	32	28	5
5	20	27	6
6	12	36	14
7	6	46	12
8	27	30	5
9	18	33	7
10	13	22	15
11	20	37	3
12	19	35	3
13	19	30	12
14	25	45	6
15	26	28	7
16	25	26	10
17	28	15	3
18	16	34	11
19	15	32	20
20	28	27	4

## **CHAPTER 4**

### **Summative assessment of vocational trainees in general practice**

#### **The results of a three year study**

##### **Introduction**

The Joint Committee on Postgraduate Training for General Practice (JCPTGP) has stated that a system of summative assessment of vocational trainees will apply to all trainees completing training after 4th September 1996 (JCPTGP, 1993). The reasons given for the decision were that the public have a right to believe that all GPs are of proven competence, and that there appeared to be evidence, if only anecdotal, that some newly qualified GPs were of doubtful competence. I described in Chapter 2 the components chosen to form the assessment process in the West of Scotland region. These comprise a multiple true/false paper, an assessment of videotaped consultations, an audit project, and the trainer's judgement. The West of Scotland video assessment instrument has been selected by the United Kingdom Regional Advisers for use on a national basis (Hasler, 1995). 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 Organiser in England) and the Trainer. Thereafter, if concern persists regarding a trainee's competence the trainee is referred for further review to a pair of assessors who make a recommendation as to whether a further period of training is required. When the system becomes nation wide this additional pair of assessors will be from a national panel of trained assessors. In the UK system the multiple true/false paper will not lead to a referral process but will produce a pass or fail result.

In Chapter 3, I have described an initial pilot study in which we assessed the reliability and feasibility of using videotaped consultations and described an instrument developed to assist assessors in forming a judgement of trainee competence. In this chapter I will describe a full scale study in which all trainees completing the trainee year between 31st July 1993 and 31st July 1995 participated in the process. The decision concerning the issuing of a certificate of satisfactory completion remained the prerogative of the trainers but it was expected that the decision would be guided by the results of summative assessment.

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

### **Subjects and methods**

Three hundred and fifty nine trainees completed the General Practice component of training during the period of the study (31st July 1993 to 31st July 1995). All trainees took part in the process and 343 took part in all components.

### **Criteria used to identify trainees for referral**

#### ***Trainer's judgement***

The trainers were asked to complete a statement giving their judgement on the trainees competence. For the first group of trainees (the July 1993 finishers) the trainers were offered a choice of three statements:

- (a) The trainee is competent to carry out the work of general practice
- (b) The trainee is of doubtful competence
- (c) The trainee is not competent and will require further training.

If the response was (b) or (c) the trainer was asked to give details of the concerns.

For later batches of trainees the trainers statement was modified as shown below. This version was developed at a workshop run by myself involving trainers from Nottinghamshire and was circulated to all West of Scotland trainers with a request for comments. All trainers found the form acceptable and suggested no major changes.

### Trainers Judgement

*Clinical competence:* Do you have any doubts about your trainees competence to perform unsupervised general practice?

yes/no

*Professionalism:* Does the 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 organisation:* Do you think the trainee will be able to cope with the organisational difficulties of general practice, particularly time management?

yes/no/doubtful

*Other areas :* Are there any other areas where you have doubts about the trainee's fitness for independent practice?

yes/no/doubtful

If yes please specify.

### *Multiple true false paper*

Each trainee sat a multiple true/false paper without negative marking. A pass mark was derived using the Angof technique as described by Livingstone and Zieky in 1982. This was done as follows: A group of experienced GP Principals analysed the paper question by question and for each question produced a figure for the percentage of trainees of minimum acceptable competence whom they would expect to answer the question correctly.

By this means minimum acceptable overall score was determined. Any candidate scoring more than one standard error below this score entered the referral process. The purpose behind using this apparently complex system was to avoid rank ordering which would result in failure of a predetermined percentage thus running 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 are as shown:

- 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 we used are those widely agreed in the literature on audit (Irvine, 1991). We did not insist on evidence of change brought about by the audit since instituting change was outwith the trainee's control.

### ***Videotape assessment***

I have discussed earlier the criteria used (Chapter 2) and the reliability of the instrument (Chapter 3) 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 entered the referral process.

### **Results**

**Trainers Judgement:** 359 trainers reports were received. Three trainees were rated as unacceptable by trainers.

**Multiple true false papers :** 358 out of 359 trainees completed these papers between March 1993 and May 1995. The missing trainee attended a different region for the day release course for geographical reasons and consequently missed the opportunity to sit the paper. As the summative assessment process developed we used MCQ papers from three different sources. 145 trainees sat a multiple true false paper developed within the region (in May 1993, December 1993, and December 1994). For the sitting in May 1994, as part of a joint initiative with the Royal College of General Practitioners, all West of Scotland trainees in the last 6 months of training sat the RCGP

MCQ, including those not taking the MRCGP exam. In March 1995 the trainees sat an MCQ based on the question bank of the fellowship exam of the Royal Australian College of General Practitioners. The Australian question bank has been shown to have excellent reliability at the pass/fail level (Hays et al, 1995). The paper was taken by over 800 trainees around the UK as part of a nationwide study. The results for the West of Scotland trainees are given here but the data from the 800 were used for the statistical analysis of the paper. The reasons for the variety of different sources are discussed later. The pre-determined pass marks were 68% (RCGP), 70% (W of S), and 76% (RACGP). The standard error was 3% (W of S) and 2% (RACGP). None of the trainees fell below the predetermined pass mark for the RCGP paper. For the locally constructed MCQ the mean score for the MCQ was 79.25% with a standard deviation of 5.46. 5 trainees scored below 70% with marks of 68% (2), 69% (2), and 64% (1). Therefore 1 trainee can be reliably identified as falling below the pass mark (95% confidence).

For the RACGP version the mean score was 75% with a standard deviation of 4.23%. 57 trainees fell below the pass mark and 20 outside the 95% confidence interval. The Alpha coefficient of the West of Scotland MCQ was calculated to be 0.76, and for the Australian based paper 0.72. It was

clear at this stage that the pass mark set for the third paper had produced a much higher fail rate than the previous papers. In view of the fact that more than half of our trainees failed to reach the pass mark we felt it inappropriate to use this pass mark and using the Hofstee technique (De Gruijter, 1985), which is discussed in detail below, the pass mark was adjusted to 70%, corresponding to the pass mark in the earlier paper. Using this pass mark no trainee fell below the 95% confidence level. The pass marks were determined using the Angof technique described earlier and the standard setting group took into account the absence of negative marking.

**Audit:** 345 audits were received. 4 were rated unsatisfactory using the criteria listed above. These trainees were given feedback and invited to re-submit. No trainee was failed as a result of the audit. 14 trainees in the first year did not submit an audit.

**Videotape assessment:** 358 videotapes were received. One trainee in the first year of the study refused to submit a tape. Each trainee videotape of consultations was viewed independently by two assessors each looking at least six consultations. The assessors were asked to identify for further assessment those trainees about whom they had doubts as to clinical

competence. Full details of this method are described in Chapter 3 and elsewhere (Campbell et al, 1995). 72 trainees were rated as refer. All of the non referred tapes were reviewed in an attempt to identify any missed referrals. One such tape was identified. One trainee was identified by both the multiple true/false paper and the videotape analysis.

### **Overall results**

There was some correlation among the four components but the majority of trainees who were identified were picked up on one component only. As shown in Table 9 a total of 77 (22%) of trainees were identified as being of potentially doubtful competence and entered the next level of the process. In order to obtain the co-operation of the region's trainers in the pilot summative assessment process it was agreed that in the first year trainers would be able to call a halt to the process in respect of their trainee at any point after the initial submission of material. 15 trainers exercised their right to halt the process at this stage and issue certificates. The three trainees deemed unsatisfactory 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 came to the conclusion that 11 were of acceptable competence and 11 were unacceptable. In an attempt to

evaluate what the results would have been for the 15 trainees who did not continue to the next phase of the process we examined in more detail those trainees where both video assessors had felt that the trainee should be referred. All of these tapes were screened by myself and four tapes, which appeared to be of unacceptable standard, were reviewed by 22 of our panel of assessors using our instrument. In the case of three of these trainees the panel was agreed that there was considerable doubt concerning the competence of the trainees.

It would appear from the above results that 17 trainees would have been refused certificates if the results had been mandatory. This would represent 5% of trainees.

The actual numbers of trainees refused certificates were 7 (3 as result of the video, 3 as a result of the trainers report, and one by both video and MCQ). The remaining 10 trainees were issued certificates by their trainers.

### **Discussion**

Although not all trainees took part in all of the process more than 95% did complete all components. We have no hard data on the non responders but

all received certificates of satisfactory completion. The external assessors agreed that in the case of all trainees who reached them there was legitimate cause for further analysis of these doctors. Analysis of the videotapes of those trainees not referred identified one case where the assessors had missed a trainee of doubtful competence. Although there was some correlation among the four components they were quite clearly identifying different problems. The trainer's report identified trainees who were disorganised, unreliable, or unable to work at an acceptable pace. The videotape identified doctors with problems in communication and patient management skills. The audit identified trainees who did not grasp the concepts of carrying out quality assurance or were unable to describe it. The MCQ identified deficiencies in knowledge.

### **The individual components**

The first two multiple choice papers produced a small number of unsatisfactory scores. If the scores within one standard error of the pass mark are excluded only 1 trainee's performance was unacceptable. It is planned to further modify the MCQ paper by introducing a larger number of problem solving and decision making tests. Setting the pass mark using the Angof technique produced on two occasions a low or non existent failure rate, on the

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 actually obtained by the candidates. A refinement of the Angof technique - the Hofstee process, may be the best way forward. In this situation as well as doing the Angof process the panel also decides acceptable maximum failure rates. Clearly this system is trying to obtain the best of both worlds and could be construed as a return to peer referencing. However, 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.

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 not felt at this stage that a refusal of certificate would be justified until further validation work had been done. The trainer's judgement produced three referrals. The UK Regional Advisers intend to use a trainers report with considerably more questions. This may produce more failures by encouraging the trainer to think more carefully about the trainee's competence.

The video component produced the majority of referrals. As discussed in our earlier papers we anticipated around 20% of trainees to be referred and between 2 and 5% to be of serious concern. This was borne out in the full scale study.

The relatively large number of trainees identified by the video who were subsequently judged to be satisfactory is, we believe, unavoidable in the interests of picking up the unsatisfactory ones. As in many situations an increase in specificity could only be achieved by a reduction in sensitivity. It would be tempting to consider for referral only those trainees identified by both assessors, however, in one case of this type our panel of assessors agreed that the trainee's performance was unsatisfactory.

An important issue is the trainer's ability to overrule the findings of the summative assessment process. Clearly summative assessment will never be credible while the trainer retains the right to issue a certificate regardless of the assessment results. A striking finding was the belief of some trainers that their trainee was competent despite the opposing evidence. It is interesting that in no such case was a trainer able to produce any objective evidence of competence such as alternative videotapes. Trainers were invited to view the

failing videotapes and in the majority of cases agreed with the factual findings of the assessors but disagreed with the need for re-training on the basis that the trainees performance was normally better than that on tape. It is worth pointing out that the trainee has the opportunity to record several tapes before submitting one which they believe to be satisfactory. Trainers are also encouraged to verify that the submitted tape is a true representation of the trainee's performance.

One of the major questions concerning summative assessment is the opportunity cost. Is it worthwhile 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 has been carried out but is outwith the scope of this thesis. For the video component, which is the most resource intensive, 2 hours of assessor time per trainee is required for the initial assessment. The national panel would look at 10% of trainee videotapes which would entail 1 days work for 40 assessors per year.

In this study the process resulted directly in 7 trainees being refused a certificate, 4 of whom had not been identified by the trainers. Our results also

suggest that a further 10 may 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 United Kingdom as a whole is unclear. It is possible that the proportion of unsatisfactory trainees in the West of Scotland is different from the rest of the country but elucidation of this will only be possible when other regions have taken part in the process.

There are those who argue that an improved system of formative assessment would render summative assessment redundant (Banks, 1994). In principle this idea has attractions but it is worth pointing out that the West of Scotland results were obtained in the presence of an extensive and mandatory formative assessment programme. In addition there remains the fundamental problem of the difficulty a trainer has in forming an unbiased view of the performance of a doctor with whom the trainer usually has a close and friendly relationship. It would require a very different relationship between trainer and trainee in order for the trainer to objectively assess consulting ability.

## **Conclusions**

The West of Scotland summative assessment programme identified a number of trainees who were not yet competent to enter general practice. The majority of unsatisfactory trainees were identified by observing their behaviour in the consultation. When the process becomes mandatory we predict that approximately 5% of trainees will be refused a certificate as a result. This has resource implications for the further training of these doctors. The results have considerable implications for patients, those involved in training, and government. Current vocational training regulations will require to be altered for summative assessment to be effective.

**TABLE 9**

	<b>Number submitted</b>	<b>Number referred</b>	<b>Number unacceptable</b>
Trainers judgement	359	3	3
Audit	345	4	0
Video Assessment	358	72	14
Multiple true/false paper	358	1	1
<b>Totals</b>		<b>77</b>	<b>17</b>

(Note - trainee who failed the MCQ also failed the video, 2 trainees who were referred as a result of the audit also were referred by the video)

## **CHAPTER 5**

### **Video recording of general practice consultations - the effect on patient satisfaction**

#### **Summary**

The study was designed to examine the effects of video recording consultations on patient satisfaction. 18 General practitioners each carried out two consulting sessions after which the patients completed a patient satisfaction questionnaire. One of the sessions was videoed, after consent had been obtained. A total of 379 completed questionnaires were obtained. Analysis of the results showed no difference in patient satisfaction between the two groups. We conclude that the presence of the video camera has no detrimental effects on patient satisfaction.

#### **Introduction**

Vidcotaping of general practice consultations has assumed a high profile recently with its proposed use in summative assessment of general practice trainees, Fellowship by assessment and the membership examination of the Royal College of General Practitioners. It has been shown by Pringle (1984,

1990) that the presence of the video camera does not alter doctor behaviour and that most patients are happy to give consent to video recording (Martin, 1984). Concern has been expressed that the presence of the camera might cause trainees to behave inappropriately and that performance would be impaired by the presence of the video camera. It is a commonly expressed view that doctors are initially conscious of the presence of the camera. Frequent use of video in formative assessment should remove any 'stage fright' and the opportunity to try again if the doctor is unhappy with the original effort should avoid excessive stress.

There has been considerable debate over the years as to the effect of the camera on patients and their vulnerability to coercion. It has been suggested (Bain, 1995) that consent rates of 4-10% are to be expected when coercion is removed. However, these studies did not involve inviting patients to be videoed. In one case (Bain, 1995) the authors asked patients to speculate on how they thought they might feel if so invited. In the second study (Servant, 1986) patients were given leaflets inviting them to volunteer if they wished to be videotaped. The proportion of patients who did not care one way or the other is included in the 90% claimed to reject videotaping. Several writers

have commented that the 10% consent rate is misleading (Boardman, 1987; Mackay, 1987; Tylee, 1987).

A study by Bain (1993) has suggested that the majority of patients would feel coerced into and uncomfortable during videotaped consultations. The concept of coercion is a difficult area to assess. Many patients will agree to take part in teaching and assessment for altruistic reasons in that they believe they will be helping to improve medical care. In a sense this could be seen as coercion. We took the view that the patients satisfaction with the consultation would be the most appropriate method of identifying any detrimental effect of the video camera. The objective our next study was to compare patient satisfaction scores after videotaped consultations and after consultations without use of video on the basis that if patients felt uncomfortable this would decrease satisfaction levels.

### **Methods and results**

A total of eighteen general practitioner trainers participated in the study. Each GP used two consulting sessions for the study. One was videoed after obtaining appropriate consent, the other was not videoed. After each consultation patients were asked to complete a consultation satisfaction

questionnaire (Baker, 1990) which has demonstrated good validity and reliability. Patients were assured of anonymity and the questionnaires were completed in the waiting room after the consultation. It has been suggested that completion in the waiting room might bias patient responses but since we were comparing groups rather than looking at overall patient satisfaction this should not affect the results. From work in a similar patient population we calculated that 100 patients per group would have a power of 90% to detect differences in satisfaction as small as 5% between the groups.

The results were analysed using SPSS-x. The results were normally distributed, and variances were homogenous using Bartlett's test. Data comparisons were carried out using Student's t-test.

379 questionnaires were returned, 182 from the video group and 197 from the non video group. The groups were well matched for age and sex. 9% (18) patients withheld consent to video recording. The findings are shown in the table. No statistically significant differences in overall satisfaction or in any of the subscales were demonstrated. Analysis by individual Practices showed no significant difference in patient satisfaction between video and non video patients for any doctor.

### Comment

The use of videotaped consultations in summative assessment would only be valid if the process did not affect the consultation in any material way. It has been suggested by Bain (1993) that the majority of patients would feel uncomfortable during a videotaped consultation. The authors of this report go on to suggest that the use of video is unacceptably intrusive. A major drawback of this study was that none of the patients involved had ever been asked to take part in a videotaped consultation.

This study demonstrates that there is no difference in patient satisfaction between a group of patients who were videoed after having given consent and a group who were not videoed. The allocation of patients to each group was random except in so far as only those patients who agreed to be videoed could be allocated to the video group. It could be argued that such patients may be different those who were not asked since the unasked group will contain some patients who would refuse to be videoed. However, over 90% of patients asked agreed to the video and in any event the ethical objections concern patients who do not refuse to take part but feel uncomfortable being videoed.

It is our assertion that if patients did feel unhappy this would be reflected in the satisfaction scores. It appears therefore that, provided appropriate informed consent is obtained consultation videotaping has no detrimental effects on patient satisfaction.

**TABLE 10**

Scale	Mean (SD) Video group Number = 182	Mean (SD) Non video group Number = 197	2 tail probability of a difference between means
Total satisfaction	72.1 (9.1)	72.2 (8.6)	.80
General satisfaction	12.6 (2.1)	12.4 (2.0)	.29
Professional care	29.2 (4.2)	29.5 (3.4)	.73
Relationship	18.9 (3.7)	18.9 (3.6)	.86
Perceived time	11.4 (2.6)	11.3 (2.5)	.58

## CHAPTER 6

### **The effects of the introduction of a system of mandatory assessment for GP trainees**

#### **Summary**

A series of questionnaires was used to determine the amount of formative assessment going on within the West of Scotland Region of the United Kingdom and its perceived value by trainees (GP registrars) both at the beginning and end of training. The first survey was carried out in 1989 and the second series in 1994. Trainers and Course organisers were surveyed in 1994 to determine their attitudes to the use of videotaped consultations for formative assessment. In 1991 the region had initiated a mandatory formative assessment programme which included regular use of videotaped consultations, confidence rating scales and Manchester ratings (RCGP Occasional Paper Number 40). The use of a range of assessment methods for formative assessment of general practice registrars (trainees) increased considerably between 1989 and 1994. The percentage of trainees using videotaped consultation analysis increased from 76% to 94%, for Manchester ratings from 52% to 68% and for confidence rating scales from 63% to 74%.

Video and confidence scales were rated highly by trainees who were exposed to them and by most trainees at the start of the year but less highly by those who had not used them by the end of the trainee year. Manchester ratings were not thought to be as useful and there was no difference here between users and non users. Despite the mandatory system a significant number of trainees (76%) were still not receiving the minimum assessment stipulated. Trainers rated lack of time as the main limiting factor to the greater use of video. We conclude that trainees who are exposed to assessment methods, particularly video, find it useful but that some trainees are still receiving less than their due. We propose increased trainer education and intensified monitoring of the assessment carried out in training Practices.

### **Introduction**

A postal questionnaire was sent to all trainees within the region completing their practice vocational training year in July 1989 (Campbell and Murray, 1990). The trainees were asked about their perceptions of the usefulness of a series of formative assessment methods and the frequency with which the assessments took place. This survey demonstrated that trainees found most of the commonly used assessment methods to be useful. However, the survey also showed that a significant minority of trainees were receiving very little

formative assessment. Since 1991, as a result of this survey, the West of Scotland region has introduced a formative assessment programme which is contained in the regional training manual held in each practice and is also distributed individually to all trainees at the beginning of the trainee year. This programme sets out the mandatory minimum assessments that training practices are to carry out. This minimum is :

Teaching on trainees videotaped recordings (on at least 3 occasions )

Use of the Manchester rating scales (2) (on at least 2 occasions)

Use of the regional confidence rating scales (on at least 2 occasions)

The Manchester rating scales are designed to enable trainers to divide the various competencies of general practice into discrete areas and to allocate scores to the trainee in each of these areas. Examples of the specific areas would be emergency care, information gathering, problem solving, and clinical judgement. The trainers are aided in this by the provision of definitions of satisfactory and unsatisfactory performance for each criterion. The method enables the trainer to identify specific areas in which the trainee

needs to improve performance. The confidence rating scale consists of a list of conditions and problems encountered in general practice. The trainee is invited to rate their confidence in dealing with these problems. This can then form the basis of a needs based educational programme.

Other methods which are encouraged at practice level are the use of written work such as essays, project work, and joint consulting sessions, although there is no statement in the programme as to how frequently these should take place.

Other assessments such as multiple choice papers and objective structured clinical examinations are also part of the programme but these are organised at district and regional level. Compliance with the regional formative assessment protocol is one of the criteria for selection and re-selection of training practices and trainers are required to state when applying for approval or re-approval that they will comply with this programme. The JCPTGP has made it a requirement that regions have such an assessment programme in place from 1st January 1993 (JCPTGP, 1992). The JCPTGP has specified that *'assessment should take place early in the post, on more than one occasion and should be based on a selection of methods'*.

### **Aims**

We set out to measure the impact of the new mandatory system on the quantity of assessment taking place in training practices, and on the trainees' perception of the usefulness of the methods used. The trainees' views on the place of the RCGP exam were also sought. In addition we attempted to explore the trainers' attitudes to the use of videotaped consultations in teaching, and on factors limiting its use.

### **Methods**

A questionnaire identical to that of 1989 was sent in June 1994 to all trainees completing training in July 1994. The trainees were advised that the responses would be analysed anonymously. The questionnaires contained an identifier for use if a second mailing to non responders was necessary. Trainees were asked to state the number of times they had been exposed to the various assessment methods. They were also asked to rate the usefulness of the various methods on a 1 to 5 ordinal scale where 1 = method is useless, to 5 = method is very useful.

In addition trainees were asked if they had been aware of an assessment programme within the practice. The opportunity was also taken to ask the trainees if they believed that the MRCGP examination was a valid assessment of competence at the end of the trainee year. Respondents were invited to supply comments in each section.

In addition a group of trainees starting their general practice year in August 1994 were asked the question regarding the usefulness of video prior to any discussion or experience of video in general practice. A questionnaire devised in the Oxford Region (Peter Havelock, 1994) was used to assess trainers' views on factors limiting the use of videotaped consultation analysis within the practice. We used the chi square test (Minitab statistical package) to compare the groups of trainees.

### **Results**

Of a total of 117 trainees, 97 responded to the 1994 questionnaire, a response rate of 83%. Fifty-four trainees completed the video question at their first regional day release meeting prior to any discussion of the use of video.

***Video consultation analysis.***

Ninety-one (94%) trainees reported that this had taken place during the year within the practice for teaching purposes. 60 (62%) reported that video assessment had taken place on at least 3 occasions during the year as prescribed in the regional package. Six trainees stated that formative video assessment had never taken place.

The mean score for all trainees on the usefulness scale was 3.69. For those who had experience of the process this rose to 3.79. For those who had never used the system the rating dropped to 2.17. The difference in scores between the groups of users and non-users was statistically highly significant (chi square,  $p < 0.001$ ). For trainees who used videotaping there was no correlation between the number of occasions video was used and the perceived usefulness. Only 2 out of 91 trainees with experience of videotaping rated it as 1 on the scale, i.e. useless, by contrast 2 out of the six trainees with no video experience rated the method as useless. For the group of 54 trainees at the beginning of the trainee year the mean score was 3.5, two trainees of this group rating video as useless. Because of the small numbers of trainees who were not exposed to video it would be inappropriate to read

too much into the statistics, but it is clear that those trainees not exposed to video rated it less highly than those who were exposed.

### *Use of Manchester ratings*

Sixty six (68%) of trainees had used Manchester ratings, and 45 (46%) had used them at least twice as specified in the Regional protocol. The average rating for the use of Manchester ratings was 3. This was the same for both users and non users.

### *Use of Confidence rating scales*

Seventy two (74%) had used these scales, with 49 (51%) using them at least twice as specified in the regional protocol. The overall average rating was 3.33, with the users giving an average score of 3.5 and the non users 2.84 ( $p < 0.01$ ).

### *Awareness of a regional programme*

Fifty (52%) trainees stated that they were aware of an assessment programme throughout the trainee year. Twenty (21%) agreed with the statement that the MRCGP examination is a valid assessment of competence at the end of the trainee year.

As stated earlier it is a requirement of the regional criteria for selection and re-selection as a training practice that the regional formative assessment programme is followed.

According to the trainee responses only 23 (24%) of practices were carrying out these minimum measures.

### *Trainers video survey*

159 responses were received from 172 trainers and course organisers, a response rate of 92%. 139 respondents had ready access to a video camera, 115 owning their own camera. The number of occasions on which trainee tapes were reviewed ranged from 0 to more than 5. When asked how often they would like to review trainee consultations 91 were happy with current usage, 51 wanted more use and 17 would choose less. Table 11 shows the details. Those trainers who felt they would like to use video more were asked to choose from a menu of factors limiting their use. Table 12 shows the results. Time was selected most commonly as the major limiting factor, followed by trainee resistance.

### *Comparisons with the 1989 survey*

In 1989 9% of trainees stated that an assessment programme was in use in the practice while 44% believed that the MRCGP was a valid assessment of competence.

Table 13 shows the number and percentages of respondents reporting the use of assessment methods. It can be seen that the use of all methods has increased considerably, although only the use of video approaches the universal use which would be consistent with a mandatory programme. Table 14 compares the opinions concerning the usefulness of the assessment methods of the two cohorts plus the opinions on video of trainees just starting the trainee year. There is in both cohorts a marked difference between users and non users for video and check list use in that users rated both these methods highly. In contrast Manchester ratings were less popular with all groups. The usefulness rating of video was unaffected by whether it was used only once or more frequently. When we compared the use of the three assessment methods with the regional protocol discussed earlier only 23 (24%) claimed to have had the mandatory minimum carried out.

## Discussion

The main findings in the earlier survey were that trainees who were exposed to the assessment methods found video recording and the use of the confidence scales useful and that trainees who used the methods had a more positive view of them than trainees who did not. This finding remains true in the current survey. Fears had been expressed that the introduction of a mandatory programme of formative assessment and the use of video in summative assessment would turn trainees against formative assessment, particularly the use of video. This has not happened since the perceived usefulness of video remained at a similar level in the second cohort who had been exposed to the use of video in summative assessment.

It is interesting that those trainees who only used video formatively on one occasion rated the video as highly as those who used it more frequently. It might have been expected that trainees would take some time to become comfortable with the presence of the video camera, and several trainees mentioned this in commenting on the use of video. Although this may be true it appears that the value of video is apparent on first use.

It is interesting that the mean score of the very small group of trainees who had not been videod was more negative in 1994 than in the earlier survey. This was typified by one non user who commented that video was artificial, threatening and unhelpful. Although the number of non users of video is very small it is possible that there is a small group of trainees to whom the idea of video is very unwelcome. This group seems to be able to avoid exposure to video, suggesting a possible degree of collusion between trainer and trainee.

An alternative explanation is that all trainees are unhappy about the use of video before they experience it and that most then find it useful once exposed to it. The evidence from trainees prior to any general practice experience with video suggests that the former explanation is correct since they had on average a positive view of video.

It is encouraging that the number of trainees exposed to the various methods has increased considerably in the 5 year period. This increase should, however, be seen in the context that formative assessment is now a national and regional requirement. It is therefore noteworthy that according to the trainees only 23 (24%) practices had completed the minimum amount of assessment specified by the regional programme. It could be argued that the

trainees were being asked to recall assessments carried out up to 11 months previously and that this would produce a bias towards under reporting. Even allowing for this it is unlikely that the true figures are dramatically higher.

We took the opportunity with both cohorts to ask the trainees their views on the appropriateness of the examination for membership of the Royal College of General Practitioners as an end point assessment for vocational training.

It is significant that at a time when the RCGP is advocating the membership examination as the entry requirement to general practice support for this has halved since the 1989 questionnaire. 80% of the cohort in 1994 were in the process of sitting the exam at the time of the questionnaire although none knew the results.

Many trainees gave reasons for their opinions. The most common reasons for rejecting the RCGP exam were on the grounds of validity - the lack of a clinical component, and fairness - the presence of a pre-determined pass rate. These factors were of course present in 1989 and do not explain the reduction in support for the MRCGP exam. It may be that the widespread debate on summative assessment in the past year has increased trainee awareness of

these aspects of the college exam and encouraged trainees to think carefully about the issues involved.

It has been argued that if formative assessment were practised thoroughly there would be no need for summative assessment (Ian Banks, National Trainee Conference, 1994; Will Copolla National Trainee Conference, 1995).

In the West of Scotland region the formative assessment programme is mandatory for training practices, and is discussed at every accreditation visit to practices. Trainees receive their copy of the assessment package and are encouraged to remind the trainer about it where necessary. Specific time is programmed into the day release programme to inform all trainees about formative assessment. Despite this less than one quarter of trainees state that they are receiving the minimum recommended assessments. This has clear implications for formative assessment within the region. The region has to date attempted to convince trainers of the educational value of formative assessment by running regular training workshops. Clearly more work needs to be done in this area to convince trainers of the value of formative assessment. There is a re-approval process which every training practice has to pass in order to continue training. During this process trainers are asked if

they comply with the regional programme and all say that they do. Although more intensive monitoring may improve the position it is unlikely that the existence of a comprehensive assessment programme within practices can ever be guaranteed.

This situation indicates that the need for some form of verified external assessment of both the competence of the trainees and the assessment processes within practices would appear to be vital since, apart from the previously discussed difficulties that trainers experience in making objective judgements about their registrars, the failure of a significant number of trainers to carry out an assessment programme would indicate that they have a poor knowledge base to even attempt this judgement.

### **Conclusions**

The introduction of a regional assessment programme has increased the amount of assessment taking place without any corresponding reduction in its perceived usefulness by trainees. A small percentage of trainees take a very negative view of the use of video and appears to successfully avoid its use. The fact that 76% of trainees still receive less than the regional minimum acceptable amount of assessment indicates that the region will have to take

further steps in trainer education and performance monitoring. Bearing in mind the current levels of in practice assessment it would be unwise to rely on this to identify unsatisfactory trainees. The attempt by the RCGP to promote the College exam as the end point assessment of training has found little favour with trainees in the West of Scotland. The introduction of a clinical component and an end to the competitive nature of the exam would answer most of the criticisms raised.

**TABLE 11**

## Trainers use of video consultations

number of occasions	actual use	ideal use
0	1	1
1	7	10
2	27	15
3	50	32
4	37	45
5	4	4
>5	17	31
<b>Total responses</b>	<b>143</b>	<b>138</b>

**TABLE 12**

## Factors limiting use of video by trainers

	a great deal					Not at all	Totals
equipment	5	8	3	2	12	36	60
time	24	15	13	6	5	5	68
low priority	0	2	19	14	12	19	68
no skill	0	2	13	10	16	24	65
trainee resistance	8	12	16	12	6	13	67

**TABLE 13**

## Trainee use of methods

	Number (percentage) 1989	Number (percentage) 1994
Video	50 (76%)	91 (94%)
Manchester ratings	33 (52%)	66 (68%)
Confidence scales	41 (63%)	72 (74%)

**TABLE 14**

## Perceived usefulness of methods to trainees

	Mean scores				
	Non-users 1989	Users 1989	Non-users 1994	Users 1994	New trainees Aug. 1994
Video	3.0	3.9	2.2	3.8	3.5
Manchester ratings	2.9	2.9	3.0	3.0	
Confidence scales	2.9	3.7	2.8	3.5	

## **CHAPTER 7**

### **The teaching and assessment of communication skills in United Kingdom Medical Schools**

#### **Introduction**

I have described in earlier chapters the development of a summative assessment programme for general practice trainees. Part of this system is the analysis of videotaped consultations. It appears from our pilot studies that this video component will be the element which identifies the highest proportion of trainees not yet ready for independent practice. In order to identify the educational background to the trainees' consulting ability it was decided to explore the undergraduate exposure to assessment of consulting skills.

#### **Summary**

All departments of General Practice in UK medical schools responded to a postal questionnaire in which we asked about the nature and extent of teaching and assessment of communication skills, the percentage of students with communication problems, the remedial teaching available and the

number of students failing to progress because of communication difficulties. In most medical schools the department of general practice was the major provider of such teaching. Estimates of students with communication problems ranged from 1% to 25% with a mean of 14% in the middle year and 12.5% in final year. Only 2 schools out of 29 claimed that some students failed to progress because of problems and in these 2 schools the percentage failing was less than 1%. Only 2 schools claimed to specifically assess communication skills in final examinations. 23 schools had plans to increase teaching and assessment of communication skills.

### **Methods and results**

A questionnaire was sent to the departments of general practice of the 29 UK medical schools. After one reminder the response rate was 100% although not all questionnaires were fully completed. We asked the following questions:

- How are communication skills taught and assessed in your medical school?
- What is the input of the department of general practice into the processes?
- What proportions of students have communication problems at the mid point of the course?

- and what proportion have problems in final year?
- Do students fail to progress as a result of poor communication skills, if so how many?
- Are communication skills assessed per se in the final MB?
- Do you have any changes planned related to the issues in the previous questions?

The results are summarised the Table 15.

The response rate of 100% is excellent. Most respondents stated that the department of general practice had a substantial input into what communication skills training was going on. A wide variety of teaching methods were used with no single method predominating. Most schools tended to use several different methods. Methods used included the Leicester Assessment package ( Fraser et al, 1994), Pendleton (Pendleton et al, 1984) and Neighbour (1992). No department felt able to give accurate numbers of students with communication problems and 7 stated that they had no idea of numbers. The wide range in numbers thought to have problems is therefore striking but unsurprising since most respondents admitted that their estimate was pure guesswork. A further remarkable finding is that communication

skills were formally assessed in the final MB exam in only two medical school and practically no students were identified as having unacceptable levels of skill.

The questionnaire encouraged respondents to make free text comments and the following are typical of the views expressed:

‘..Even students with serious problems appear to get to finals undetected’,  
‘failure to assess communication skills in finals is a major deficit’,  
‘little or no interest shown by our hospital colleagues’.

### **Discussion**

The General Medical Council as long ago as 1980 (General Medical Council) recommended that doctors should be able .... “to communicate effectively and sensitively with patients and their relatives”. More recently in 1993 communication skills were one of the eight curriculum themes laid down by the GMC (GMC: Tomorrow’s Doctors). In our work in the summative assessment of vocational trainees in general practice (Campbell and Murray, 1996) around 5% of trainees display consulting skills which do not attain minimum acceptable competence.

It is not only the medical profession which sets store by the possession of consulting skills. The public rates the ability of a doctor to listen as a high priority. Many studies have shown that good consulting and listening skills not only improve patient satisfaction (Grol et al, 1990; Comstock et al, 1982) but also improve outcome (Bass et al, 1986; Henbest et al, 1990) and make it easier to alter patient behaviour (Arborelius et al, 1992). However, it is also important that doctors develop a flexible consulting style since there is evidence that patients with physical problems prefer a directive style of consulting (Savage and Armstrong, 1990).

It is therefore encouraging that 26 universities had a programme of communication skills teaching. The fact that departments of general practice in most cases took the lead role is appropriate given the level of expertise and experience general practice departments have developed. It is, however, extremely disappointing that this expertise is not being used to diagnose and deal with communication problems. Although attempts are being made to teach communication skills it is surprising that little attempt is made to assess the results of this teaching. What evidence there is suggests that communication skills teaching as presently practised has little effect on

students' behaviour (Moorhead and Winfield, 1991). Assessment of the effectiveness of teaching is particularly important since many departments use methods of teaching developed "in house" and there is clearly no agreed national curriculum. Communication skills must be the only area of the medical curriculum where students are taught but are neither assessed nor failed. What would we think of a medical school which did not assess the knowledge base of its students? It is also surprising that practically no students failed to progress as a result of detected poor communication skills despite around 10% being identified as having poor skills. This 10% is certainly an underestimate since there is no system in place in most medical schools to detect the problem. The GMC said in 1993 that 'deficiencies in this area are responsible for a high proportion of complaints and misunderstandings'.

It is encouraging that most departments expressed strong views that communication skills will, with the advent of new curricula, assume a much higher profile. Hopefully this will prevent the situation which obtains currently where communication problems are identified at a much later stage in the doctor's career. In the light of the evidence presented above it is certainly not surprising that our system of summative assessment does

identify a significant number of doctors who are seriously deficient in consulting skills.

TABLE 15

## Department involvement

Department involved	Percentage with		Remedial work done	Percentage who fail	Assessed in finals	Plans to do more
	Problems mid-point	Finals				
yes	10	not known	no	0	no	yes
no	20	20	no	0.25	no	yes
yes	25	25	no	0	no	yes
yes	20	20	no	0	no	yes
yes	5	2	no	2	yes	yes
yes	not known	not known	no	0	no	no
yes	not known	not known	yes	2	no	yes
yes	10	10	no	0	no	yes
yes	1	1	yes	0.2	yes	yes
yes	not known	not known	no	0	no	yes
yes	5	5	no	0	no	no
yes	25	25	yes	0	no	yes
yes	20	20	no	0	no	yes
yes	7.5	7.5	no	0	no	yes
yes	not known	not known	no	0	no	yes
yes	not known	not known	no	0	no	yes
yes	not known	not known	no	0	no	yes
yes	not known	not known	yes	no response		
yes	not known	not known	yes	0	no	yes
no	1.5	3.5	yes	0	no	no
yes	20	20	no	0	no	yes
yes	10	10	no	0	no	yes
yes	not known	not known	yes	0	no	yes
yes	not known	not known	no	no response		
yes	2	2	no	0	no	yes
yes	15	10	yes	0	no	yes
yes	not known	not known	yes	0	no	yes
yes	20	20	no	0	no	no
yes	don't know	don't know	no	0	no	yes

## CHAPTER 8

### Discussion

The null hypothesis for this thesis was that the system of summative assessment in place at the commencement of our work was effective. The analysis of the existing process demonstrated that there was a significant shortfall in the quantity and quality of assessment taking place within training practices, at least in the West of Scotland region.

In developing a reliable, valid and feasible assessment package we have recognised several areas that required to be addressed:

1. The system should be criterion referenced.
2. Several discrete assessment methods would be needed to successfully differentiate between the competent and the non competent.
3. Each method should be independently valid and reliable.

In developing the assessment package we have shown that a reliable and valid assessment of consulting skills can be carried out by trained assessors in a feasible time scale.

In terms of outcome there has been a considerable increase in the number of GP registrars recommended for additional training. The number so identified in the West of Scotland has increased markedly and if the figures are repeated across the United Kingdom there will be a tenfold increase in doctors being refused certificates of satisfactory completion as a direct result of the introduction of summative assessment.

One of the major changes which has resulted from the implementation of the package is that in the fields of knowledge, consulting and audit there is now assessment from outwith the training practice and in the area of the trainers report the trainer, although retaining the right to make a judgement about the GP registrar, now has to produce written evidence of the grounds for his/her decision.

#### ***The development of the National Panels for video and audit***

Although the system as developed in the West of Scotland used external assessors this was done on an ad hoc basis. The UK Regional Advisers have now organised a national panel of assessors for the video and audit submissions. The members of the two national panels have received initial training from the West of Scotland and will undergo periodic training and

calibration. The distribution of tapes and audits to national panel members is now organised centrally.

### *Quality assurance of videotape assessment*

In order to satisfy the JCPTGP that the consultation assessment is satisfactory the following procedures have been put in place.

#### A. Within the Regions

##### 1. *Adequate sensitivity*

The smallest feasible number of unsatisfactory trainees should be missed. The West of Scotland data suggest that around 4% of trainees will be unsatisfactory and 5% of the unsatisfactory trainees will not be identified. In numerical terms for every 1000 trainees going through the system 40 will be unsatisfactory of whom we would expect to miss 2. It is hoped that this level will satisfy the JCPTGP bearing in mind that any attempt to improve on this figure would necessitate a large increase in the number of trainees being scrutinised in detail and a huge increase in workload. In order to maintain this level of sensitivity the Regions will carry out the following:

- i) group calibration sessions looking at borderline tapes from the most recent batch of trainees, six monthly at first, thereafter on an annual basis.
- ii) internal quality control by assessing a random selection of tapes which did not reach the level of more detailed scrutiny. Regions will carry out this process on at least 20% of non referred tapes, with a reduction to 10% if the results justify this after the first two years.
- iii) identification of assessors whose judgements are inconsistent with their peers with appropriate remedial action.

## 2. *Adequate specificity*

It is very unlikely that an acceptable trainee would be refused a certificate since the system has an inbuilt series of reviews. The major component in this system is the presence of an external element in the later stages.

### B. Externally

#### 1. *Sensitivity*

It is vital that the process is fair and is seen to be fair. A major element of this fairness is that identical conclusions should be reached about trainees of

similar competence no matter which region carries out the assessments. To this end all tapes identified at the first stage of video assessment as being potentially unsatisfactory will be reviewed by a pair of second level assessors who will make a final recommendation. These external assessors will take part in some of the training and review work of regional panels but will not be involved in the first level assessment.

## 2. *Comparability of regions*

It is important that decisions are consistent throughout the regions. To this end the UK Conference of Regional Advisers will take responsibility for monitoring regional standards.

There will be 3 different categories of results produced with corresponding monitoring requirements. The monitoring system will not affect the results of individual trainees.

- 1) Tapes passed by both first level assessors (80%, i.e. approximately 1600 per year).
- 2) Tapes entering referral process but ultimately passed (16%, i.e. 320 per year).

- 3) Tapes of trainees who fail the process (approximately 4%, i.e. 80 per year).

The initial global quality assurance will consist of collating the results and confirming that categories 1 and 2 are in the appropriate proportions. The proportion of category 3 will also be recorded but there will inevitably be considerable variation since the number of non competent trainees is very small but liable to considerable variation.

For category 1, it will be necessary to review a sample of these to confirm that regions are not allowing unsatisfactory tapes through at this stage. A 1 in 20 sample should give adequate data and would involve one central assessor to view each tape for approximately one hour. This will entail a total of 80 assessor hours.

For category 2 there is a higher rate of sampling since this group will contain trainees on the borderline where decisions will be most difficult and where any regional variation will have considerable effect on the overall results. A sampling rate of 1 in 5 is proposed, i.e. 64 tapes per year. This will require 64 assessor hours.

Category 3 (failed trainees) should be scrutinised intensively, at least initially, since it is here that it is vital that regional variation is minimal. This will also be the smallest category. It is proposed that initially all such tapes are reviewed centrally (80 assessor hours). It must be emphasised that this review will not attempt to alter decisions already made but will be used to guide regional teams in future assessments.

The random sampling will be achieved by allocating every trainee in the each region a unique number and identifying those for review using appropriate random number generation. The central assessors will be blinded for the results of the regional decisions i.e. they will not know if they are dealing with a failed trainee or not.

### *Costings*

There will be some secretarial time at regional and central level plus the transportation costs for the tapes. The results of the local Quality Assurance will require to be collated and reported to the region and the UK Conference of Regional Advisers. This will require some Adviser time. The central QA will also require collation and analysis.

The main central cost will be the time of the central panel of assessors. This will be 224 times £X, where £X = the hourly rate paid to the central assessors.

At the time of writing it appears that funding will be made available to the UK Regional Advisers to carry out this work.

### **The effect of the introduction of the package on patients and GP Registrars**

Those speaking for GP registrars have been consistent in opposing the introduction of summative assessment. At three successive UK National Trainee Conferences there have been votes rejecting summative assessment and the trainee subcommittee of the General Medical Services Committee has remained resolutely opposed to summative assessment. It is worth bearing in mind, however, that a tiny proportion of trainees actually attend the national conferences and that the 'leadership' of the GP registrar sub committee may not represent the views of the majority of trainees. However, it is quite likely that the average GP registrar is not wholly in favour of summative assessment. There has been anecdotal evidence put forward that summative assessment increases the stress levels of registrars although a survey, as yet

unpublished, carried out by the GP registrars in the West of Scotland, showed that the RCGP exam was a more significant source of stress than summative assessment.

A booklet has now been issued by the UK Regional Advisers, enclosed, giving guidance to GP registrars on the summative assessment process.

Much has been said about the effects on patients of the video component of summative assessment. It is interesting that videotaping of consultations had been taking place with very little fuss until it became high profile in the summative assessment context. Discussion of videotaped consultations has now appeared in *The Observer*, *Sainsbury's* magazine, the *Glasgow Evening Times* and the *Ayr Advertiser* and may well have appeared in other non indexed journals.

Useful outcomes of this increased publicity were that we, as a result, carried out the patient satisfaction study discussed earlier. In addition the General Medical Council carried out an investigation into videotaping and came to the conclusion that it was acceptable, provided appropriate informed consent was obtained.

The author and colleagues met with Local Health Councils and with the Scottish Association of Health Councils. These bodies, while having concern for patient confidentiality, supported the aims of summative assessment.

### **The future role of the summative assessment system**

At the time of writing the complete system is in place and more than three hundred trainees, completing training after September 1996, have taken the MCQ component of the process. In the autumn of 1996 significant numbers from outwith the West of Scotland will be taking the other components. All systems are now in place for this. It is anticipated that the training regulations will change in 1997 to make satisfactory performance in summative assessment a requirement for certification. Between September 1996 and the change in the regulations it has been agreed by the JCPTGP and the UK Regional Advisers that the process will be professionally led and that trainers will be 'guided' by the results. The JCPTGP has decided that it will review all of the available evidence before issuing a certificate of satisfactory completion and that it will only issue such a certificate if it is satisfied that the doctor is competent. There remains, during this transition period, the possibility that a GP Trainer will decide to act contrary to the results of

summative assessment by issuing a certificate of satisfactory completion. The Joint Committee on Postgraduate Training for General Practice, as the competent authority, would then have to weigh up the available evidence before coming to a decision.

### **The relationship between summative assessment and the RCGP exam**

The UK Regional Advisers and the RCGP have both stated that all doctors entering general practice should pass the examination for membership of the RCGP. However, there is as yet no evidence that the RCGP examination is a reliable assessment of minimum competence. From the autumn of 1996 the RCGP examination will contain an assessment of consulting skills by analysis of submitted video recorded consultations. If this assessment can be shown to be as reliable as the summative assessment video component in identifying the non competent doctor there will be a good case for this to exempt candidates from the need to submit a summative assessment video. The RCGP multiple choice paper has already been accepted as an exemption from the summative assessment MCQ. There are as yet no proposals to provide alternatives to the UK Regional Advisers' trainer's report or the submission of practical work.

### **The effects of summative assessment on medical education**

It has long been recognised that an assessment process has a major influence on the curriculum. Whenever there are changes in assessment the curriculum is usually very quick to adapt. This has already been the case as far as summative assessment is concerned. A course on consulting skills teaching in England has recently advertised that it will help trainers prepare their trainees for the summative assessment consulting skills component. It is highly likely that there will be an increasing emphasis on consulting skills, knowledge and quality assurance in vocational training programmes in years to come. One striking development in the West of Scotland over the past three years has been an increase in the proportion of training practices possessing their own video camera. This has increased from 25% in 1993 to 90% in 1996. Although video cameras are cheaper and easier to use than in the past it seems highly likely that the regional requirements for video use have stimulated this development. Many practices are using their cameras for other training related activities such as analysis of tutorials and review of partners' consultations. The summative assessment programme has also been instrumental in identifying training practices where formative assessment has been inadequate. Clearly trainers cannot be held responsible for the lack of competence of a trainee but there is a presumption that a trainer should be

able to identify the doubtful trainee at a relatively early stage. In several instances we have been able to identify situations in which trainers have been remiss in failing to identify unsatisfactory trainees and appropriate action has been taken. This will hopefully have an influence on all trainers.

As we have shown earlier the teaching and assessment of communication skills at the undergraduate level is not well advanced at the moment but this is changing rapidly as universities adopt curriculum changes which will place an increased emphasis on utilisation rather than possession of knowledge.

### **Overall conclusions**

The aim of this thesis was firstly to evaluate the effectiveness of the pre-existing system for the summative assessment of doctors at the end of vocational training. The results show that there were significant deficiencies in the pre-existing system.

The second objective was to investigate the possibility of introducing reliable external assessments of consulting, knowledge and audit. Methods of acceptable reliability were developed which produced a greatly increased pick-up rate of unsatisfactory performance.

Because of concerns expressed about the effect of videotaping on patients we then carried out a study to look at the effects of videotaping on patient satisfaction. This demonstrated that the use of the video camera had no effect on patient satisfaction.

As a result of the significant number of doctors identified as having poor consulting skills we carried out a survey of UK medical schools to identify the extent of teaching and assessment of communication skills to undergraduates. The results showed that teaching was limited and assessment virtually non-existent.

Finally, in Chapter 8, I have described the process of converting the results of research into the reality of a UK wide system of summative assessment. The summative assessment programme was developed from first principles. We then carried out studies locally to validate the system and introduced it within the region despite opposition locally from some trainers and trainees. The arguments of those opposed to summative assessment were addressed and dealt with. The programme was then put forward as a national system. Despite considerable opposition from the Trainee Subcommittee of the

GMSC and elements within the JCPTGP the consulting skills, audit and MCQ components have been approved by the JCPTGP and the UK Regional Advisers. The video and audit components are currently the only assessments of consulting skills and written submission of practical work to be accepted by the Joint Committee. The video component was reviewed by the UK Regional Advisers along with three other consulting assessment models and was agreed to be the preferred option.

Although the system has been adopted by the UK Regional Advisers the author remains responsible for running the MCQ component and for the implementation of the consulting skills element. Five training courses for video assessors have been held by our department throughout the UK and more than 300 assessors have received initial training. The author will be responsible for continuing training of the national panel.

The implications of this work will hopefully spread beyond initial training for general practice. There is now a widespread acceptance that it is possible to assess the performance of GPs by the observation of real consultations on videotape. This raises the possibility that recertification could eventually involve more than attendance at postgraduate meetings and could one day

involve periodic direct assessment of performance. This has considerable implications for general practice. A change in the political climate would be needed, however, before such a system gained acceptance by general practitioners as a group.

It is hoped that the development of the summative assessment will continue and that the system described in this thesis will be first, but by no means the final word in the development of summative assessment of vocational trainees.

## ACKNOWLEDGEMENTS

I would like to thank my colleagues in the Glasgow Department of Postgraduate Medical Education for their advice, support and help with video assessor training: Diane Kelly, Moya Kelly, George Dyker and Murray Lough. Others who helped with video training were Jim Herbert, Colin Hodgson and David Blair. Thanks are also due to administrative and secretarial staff in the postgraduate office, headed by Isabel Robertson, for their help both with the production of the thesis and in the running of summative assessment.

co-authors of aspects of the work were Professor John Howie, Professor Stuart Murray and Frank Sullivan to whom I am indebted. I would also like to thank Peter Tate and Richard Wakeford of the RCGP for advice at the beginning of the project and lively debate throughout.

I owe a deep debt of gratitude to Professor Stuart Murray without whose continual encouragement I would never have completed the project. Finally I must thank my wife Pat and the children, Louise, Joanna and Stuart for their forbearance.

### List of graphs and tables

		<b>Page</b>
Table 1	Recorded use of assessment methods	46
Figure 1	Assessments reported by trainers	47
Figure 2	Assessments reported by trainees	47
Table 2	Trainers' ratings of assessment methods	48
Table 3	Trainees' ratings of assessment methods	48
Table 4	Responses to standard assessment programmes	49
Figure 3	The effect of passing or failing a component	66
Table 5	Correlations among sub-scales	99
Table 6	Assessor marking behaviour	99
Table 7	Overall assessor decisions	100
Table 8	Perceived challenge of consultations	100
Table 9	Overall results, the first three years	120
Table 10	Patient satisfaction video v non video	127
Table 11	Trainers' use of video	144
Table 12	Factors limiting use of video by trainers	144
Table 13	Trainees' use of assessment methods	145
Table 14	Perceived usefulness of assessment methods	145
Table 15	Communication skills teaching in medical schools	153

## **List of Appendices**

Appendix A	Original Publications
Appendix B	Consultation log
Appendix C	Video assessment marking schedule
Appendix D	Video assessment instructions
Appendix E	Current video consent form
Appendix F	UK Regional Advisers - instructions to trainees

## Appendix A

### Original Publications

Campbell L M, Murray T S. 1990. *Trainee assessment - a regional survey*. British Journal of General Practice. 40 (341): p507-509.

Campbell L M, Howie J G R, Murray T S. 1993. *Summative assessment : A pilot project in the West of Scotland*. British Journal of General Practice. 43: p430-434.

Campbell L M, Murray T S. 1993 *General practitioners' performance {letter}*. British Medical Journal. 306 (6889): p1417.

Campbell L M, Murray T S. 1994. *Leicester Assessment Package {letter}*. British Journal of General Practice. 44 (388): p535.

Campbell L M, Murray T S. 1995. *Dealing with Duffers {letter}*. Lancet. 346: p1164.

Campbell L M, Sullivan F, Murray T S. 1995. *Videotaping of general practice consultations : effect on patient satisfaction*. British Medical Journal. 311: p236.

Campbell L M, Howie J G R, Murray T S. 1995c. *The use of videotaped consultations in the assessment of general practice trainees*. British Journal of General Practice. 43: p430-434.

Campbell L M, Murray T S. 1995. *Reliability of the MRCGP viva. {letter}* British Medical Journal. 311 (7019): p1572-1573.

Campbell L M, Murray T S. 1995. *Videotaped consultations {letter}*. British Journal of General Practice. 45 (400): p631.

Campbell L M, Murray T S, Sullivan F M. 1996. *Videotaping of general practice consultations*. British Medical Journal. 312 (7025): p248.

Campbell L M, Murray T S. 1996. *The effects of the introduction of a system of mandatory formative assessment for general practice trainees*. Medical Education. 30: p60-64.

Campbell L M, Murray T S. 1996. *Summative assessment in the West of Scotland - the results of the first 3 years*. British Journal of General Practice. 46(408): 411-414

Campbell L M, Murray T S. 1996 *The assessment of competence*. British Journal of General Practice. 46(411): 619-622

## Appendix B

### CONFERENCE OF POSTGRADUATE ADVISERS IN GENERAL PRACTICE

#### UNIVERSITIES OF THE UNITED KINGDOM

#### VIDEOTAPE LOG

Trainee name .....

Consultation number .....

Camera clock time .....

Reason for patient's attendance .....

.....

Patient's age ..... Male/Female .....

Physical findings, if any .....

Action taken, e.g. prescription .....

In approximately 50 words please outline the setting of the consultation  
and what was achieved, what issues may arise later?

Please rate the degree of difficulty of this consultation by circling the  
appropriate response

straightforward/moderate/difficult

## Appendix C

### WEST OF SCOTLAND SUMMATIVE ASSESSMENT PROCEDURES FOR GP TRAINEES

#### VIDEO ASSESSMENT PROTOCOL, INSTRUCTIONS TO ASSESSORS

Attached is a "Video Assessment Form" which is an instrument which has been designed to help GP Assessors to make fair, valid and reliable judgements about GP Trainees.

These first two sheets indicate the recommended way of selecting and looking at consultations which have been recorded on video and provided by Trainees. For each Trainee, you should have in addition to a video, a booklet providing the following information:

- (a) a log of all the consultations on the tape, with time/date or index numbers; and
- (b) an assessment form for each consultation, recording Trainees' brief comments.

Your tasks are as follows :

1. Select 10 consultations for *possible* viewing. These should cover the spectrum of age groups and consultation types. Most should offer more than a low degree of challenge.
2. Make a note of each selected consultation's number on the tape (e.g. 1, 3, 5, etc) and if possible its start time (as indicated by any on-picture clock).
3. Examine the information available to you about the first selected consultation. Now view the consultation (*in silence*, if you are looking at it with a colleague). On the "Video Assessment Form", this is "Consultation 1"; record the number of the consultation on the tape and the topic covered. As the consultation proceeds, note briefly any strengths and weaknesses - and any mistakes - in the boxes provided.
4. When the consultation ends, review the relevant notes made on it by the Trainee. Add any further comments under "strengths" and "weaknesses".

Then rate the following aspects of the consultation *without comment or discussion*. But beware of making inferences from clinical presentations which you did not see (e.g. a rash).

\* First, if there was/were evident diagnostic, management or other errors made, check one of the "error/s" boxes on the form. A major error is one which causes actual or potential harm, a minor error causes inconvenience only.

\* Now try to ignore such errors. Please rate the trainee on the three criteria of "listening", "action" and "understanding" on the following scale :

1	Refer
2	Probably refer
3	Bare pass
4	Competent
5	Good
6	Excellent

Listening :

The trainee should identify and elucidate the reason/s for the patient's attendance. A credible and mutually acceptable plan should be negotiated with the patient.

Action :

The trainee should take appropriate action to identify the patient's problem/s. Investigations and referral should be reasonable. Help should be sought when necessary. The patient's problem/s should be managed appropriately.

Understanding :

The trainee should demonstrate in the workbook that he/she understood the process and outcome of the consultation. Individual actions should be explained. Obvious shortcomings in the consultations should be identified and relevant background should be identified.

The scales are designed to help you towards your judgement of this consultation and to enable you to record detailed opinions as evidence for your decisions on this consultation and, later, overall.

Tick the relevant number in the spaces provided on this part of the Assessment Form.

5. Next, note the difficulty or extent of the challenge the consultation provided - for end-point of training, was this low, medium or high? Tick one of the boxes.
6. With the information now before you on the form, judge the consultation PASS/REFER for this stage of training (i.e. completion of it). Tick "P" or "R" to indicate your judgement, on balance. But if the consultation was either so well or so badly managed for you to be completely clear and sure about your judgement, tick "P+" or "R+" instead.
7. Now repeat steps 3 - 6 for subsequent consultations. It is essential that you view and rate at least six consultations for a candidate. If at this point the consultations all point to a clear outcome or you feel that you can make a confident overall judgement, go to Step 8. Otherwise, view more consultations as required (to a maximum of 10).
8. Turn to the bottom of the last page of the Assessment Form. Summate your judgements (i.e. how many "P+s", etc), and indicate your overall judgement (which *must* be "P" or "R"). Bear in mind that "refer" may merely mean that you have doubts about a trainee. Therefore if for any reason you cannot reach a firm decision, the outcome must be "refer". Justify your decision briefly in the space provided.

Make sure that your own and the Trainee's name are clearly indicated, and that the form is signed and dated.

# West of Scotland Summative Assessment Procedures for General Practitioner Trainees VIDEO ASSESSMENT FORM

CONSULTATION	STRENGTHS	WEAKNESSES	ASSESSMENT
*1 Topic:			* Major error/s? [ ] (identify in * Minor error/s? [ ] prev. column) * Listening _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] _____ * Action _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] _____ * Understanding _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] _____ * Challenge _____ low [ ] medium [ ] high [ ] _____ * ASSESSMENT _____ R+ [ ] R [ ] P [ ] P+ [ ]
Comments on Trainee's notes about consultation			
No. on Tape:			
Time of Start:			

CRITERIA		Rating Scale	Overall Assessment
Error/s	The presence of a single major error on the consultation or of a number of minor errors should lead to consideration of referral. (serious error = causes actual/potential harm; minor error = inconvenience only)	1 Refer 2 Prob.referr 3 Bare Pass 4 Competent 5 Good 6 Excellent	R+ Clear refer R Refer P Pass P+ Clear pass
Listening	Identify and elucidate reason/s for attendance. A credible/acceptable plan should be negotiated.		
Action	Approp. action to i/d patient's problems. Reasonable investigations/referrals. Help sought when necessary. Patient's problems should be managed appropriately.		
Understanding	Trainee understands process/outcome of consultation. Actions explained. Obvious shortcomings identified and relevant background mentioned.		

CONSULTATION	STRENGTHS	WEAKNESSES	ASSESSMENT
*2 Topic:			* Major error/s? [ ] (identify in * Minor error/s? [ ] prev. column) * Listening _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] _____ * Action _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] _____ * Understanding _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] _____ * Challenge _____ low [ ] medium [ ] high [ ] _____ * ASSESSMENT _____ R+ [ ] R [ ] P [ ] P+ [ ]
Comments on Trainee's notes about consultation			
No. on Tape:			
Time of Start:			

CONSULTATION	STRENGTHS	WEAKNESSES	ASSESSMENT
* 3 Topic:			* Major error/s? [ ] (identify in * Minor error/s? [ ] prev. column) * Listening _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] _____ * Action _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] _____ * Understanding _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] _____ * Challenge _____ low [ ] medium [ ] high [ ] _____ * ASSESSMENT _____ R+ [ ] R [ ] P [ ] P+ [ ]
No. on Tape:	Comments on Trainee's notes about consultation		
Time of Start:			

CRITERIA	
Error/s	The presence of a single major error on the consultation or of a number of minor errors should lead to consideration of referral. (serious error = causes actual/potential harm; minor error = inconvenience only)
Listening	Identify and elucidate reason/s for attendance. A credible/acceptable plan should be negotiated.
Action	Approp. action to i/d patient's problems. Reasonable investigations/referrals. Help sought when necessary. Patient's problems should be managed appropriately.
Understanding	Trainee understands process/outcome of consultation. Actions explained. Obvious shortcomings identified and relevant background mentioned.

Rating Scale	Overall Assessment
1 Refer	R+ Clear refer
2 Prob.referr	R Refer
3 Bare Pass	P Pass
4 Competent	P+ Clear pass
5 Good	
6 Excellent	

CONSULTATION	STRENGTHS	WEAKNESSES	ASSESSMENT
* 4 Topic:			* Major error/s? [ ] (identify in * Minor error/s? [ ] prev. column) * Listening _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] _____ * Action _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] _____ * Understanding _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] _____ * Challenge _____ low [ ] medium [ ] high [ ] _____ * ASSESSMENT _____ R+ [ ] R [ ] P [ ] P+ [ ]
No. on Tape:	Comments on Trainee's notes about consultation		
Time of Start:			

CONSULTATION	STRENGTHS	WEAKNESSES	ASSESSMENT
* 5 Topic:			* Major error/s? [ ] (identify in * Minor error/s? [ ] prev. column) * Listening _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] _____ * Action _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] _____ * Understanding _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] _____ * Challenge _____ low [ ] medium [ ] high [ ] _____ * ASSESSMENT _____ R+ [ ] R [ ] P [ ] P+ [ ]
No. on Tape:	Comments on Trainee's notes about consultation		
Time of Start:			

CONSULTATION	STRENGTHS	WEAKNESSES	ASSESSMENT
* 6 Topic:			* Major error/s? [ ] (identify in * Minor error/s? [ ] prev. column) * Listening _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] <hr/> * Action _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] <hr/> * Understanding _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] <hr/> * Challenge _____ low [ ] medium [ ] high [ ] <hr/> * ASSESSMENT _____ R+ [ ] R [ ] P [ ] P+ [ ]
Comments on Trainee's notes about consultation			
No. on Tape:			
Time of Start:			

CRITERIA	Rating Scale	Overall Assessment
Error/s      The presence of a single major error on the consultation or of a number of minor errors should lead to consideration of referral. (serious error = causes actual/potential harm; minor error = inconvenience only)	1    Refer 2    Prob.refere 3    Bare Pass 4    Competent 5    Good 6    Excellent	R+    Clear refer R      Refer P      Pass P+    Clear pass
Listening      Identify and elucidate reason/s for attendance. A credible/acceptable plan should be negotiated.		
Action        Approp. action to l/d patient's problems. Reasonable investigations/referrals. Help sought when necessary. Patient's problems should be managed appropriately.		
Understanding    Trainee understands process/outcome of consultation. Actions explained. Obvious shortcomings identified and relevant background mentioned.		

CONSULTATION	STRENGTHS	WEAKNESSES	ASSESSMENT
* 7 Topic:			* Major error/s? [ ] (identify in * Minor error/s? [ ] prev. column) * Listening _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] <hr/> * Action _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] <hr/> * Understanding _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] <hr/> * Challenge _____ low [ ] medium [ ] high [ ] <hr/> * ASSESSMENT _____ R+ [ ] R [ ] P [ ] P+ [ ]
Comments on Trainee's notes about consultation			
No. on Tape:			
Time of Start:			

CONSULTATION	STRENGTHS	WEAKNESSES	ASSESSMENT
* 8 Topic:			* Major error/s? [ ] (identify in * Minor error/s? [ ] prev. column) * Listening _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] <hr/> * Action _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] <hr/> * Understanding _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] <hr/> * Challenge _____ low [ ] medium [ ] high [ ] <hr/> * ASSESSMENT _____ R+ [ ] R [ ] P [ ] P+ [ ]
Comments on Trainee's notes about consultation			
No. on Tape:			
Time of Start:			

CONSULTATION	STRENGTHS	WEAKNESSES	ASSESSMENT
* 9 Topic:			* Major error/s? [ ] (identify in * Minor error/s? [ ] prev. column) * Listening _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] <hr/> * Action _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] <hr/> * Understanding _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] <hr/> * Challenge _____ low [ ] medium [ ] high [ ] <hr/> * ASSESSMENT _____ R+ [ ] R [ ] P [ ] P+ [ ]
No. on Tape:		Comments on Trainee's notes about consultation	
Time of Start:			

CRITERIA	Rating Scale	Overall Assessment
<b>Error/s</b> The presence of a single major error on the consultation or of a number of minor errors should lead to consideration of referral. (serious error = causes actual/potential harm; minor error = inconvenience only)	1 Refer	R+ Clear refer
	2 Prob.referr	R Refer
	3 Bare Pass	P Pass
	4 Competent	P+ Clear pass
	5 Good	
	6 Excellent	
<b>Listening</b> Identify and elucidate reason/s for attendance. A credible/acceptable plan should be negotiated.		
<b>Action</b> Approp. action to i/d patient's problems. Reasonable investigations/referrals. Help sought when necessary. Patient's problems should be managed appropriately.		
<b>Understanding</b> Trainee understands process/outcome of consultation. Actions explained. Obvious shortcomings identified and relevant background mentioned.		

CONSULTATION	STRENGTHS	WEAKNESSES	ASSESSMENT
* 10 Topic:			* Major error/s? [ ] (identify in * Minor error/s? [ ] prev. column) * Listening _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] <hr/> * Action _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] <hr/> * Understanding _____ 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] <hr/> * Challenge _____ low [ ] medium [ ] high [ ] <hr/> * ASSESSMENT _____ R+ [ ] R [ ] P [ ] P+ [ ]
No. on Tape:		Comments on Trainee's notes about consultation	
Time of Start:			

**SUMMATION** Numbers of consultations rated: R+ \_\_\_\_\_

**Overall Decision**

Refer [ ]

Pass [ ]

R \_\_\_\_\_

P \_\_\_\_\_

P+ \_\_\_\_\_

**Comments : Reasons for Refer Decision**

.....

.....

Trainee ..... Assessor ..... Date ..... Signed .....

## Appendix D

### VIDEO ASSESSMENT - INSTRUCTIONS

As you know the region has a programme of summative assessment and I enclose the material for your videotape. It is important that you follow these instructions which are designed to expedite the work of the Assessors. The tapes will be viewed only by the assessors who are all GPs and by educational advisers working with the assessors.

The tape should be of 2 hours duration in standard VHS format. Do not use long play. If you have a camera which uses any other format this must be transcribed into standard VHS. All consultations should be recorded except where the patient withholds consent. Consent should be obtained in writing. We enclose a consent form which we believe to be suitable. We strongly suggest that you do not delete consultations which appear unsatisfactory. All of us have consultations which do not go well. You need not switch off the camera between consultations unless the gap is likely to be a long one. The examination couch should not be in shot, and examinations should not be filmed but you must keep the tape running while examination takes place since the consultation usually continues during the examination. Incomplete consultations **must** be deleted. Please check before sending the tape in that the quality of sound and picture is adequate. A desk microphone is very useful in obtaining good sound quality. Camera microphones tend to pick up lots of extraneous noise. Consultations must be of realistic length. The average time for a consultation should not be more than 15 minutes.

A camera clock is important since it allows the Assessor to navigate through the tape. If your camera does not have a clock a possible alternative is to have a clock in view on the desk. The tape counter is useless since these vary from recorder to recorder.

Please complete a log entry for each consultation. The time to be noted for each consultation is that of the camera clock. Make sure that this is switched on throughout the recording. You are asked to comment on each consultation. It will therefore be necessary for you to view the tape yourself. This will also provide a check on the quality of reproduction. Your understanding of the strengths and weaknesses of each consultation is something the Assessors will be looking at. It is therefore important that you take some care in completing the log.

If the technical quality is not good enough you will have to do another tape. Each tape and log page should be clearly identified with your name and address. Completed tapes and logs should be returned by the given date to the address below. Please either use registered post or hand deliver the tapes yourself. If you have any queries do not hesitate to contact me at the address below.

**Dr Malcolm Campbell  
Assistant Adviser in Assessment  
Department of Postgraduate Medicine  
University of Glasgow  
1 Horselethill Road  
Glasgow G12 9LX**

Thank you for your co-operation.

You must demonstrate in your video tape that you are able to:

- identify the reasons for the patients attendance.
- take appropriate steps to investigate the problems presented.
- organise a suitable management plan.
- reach an agreement with the patient on diagnosis and treatment.
- demonstrate an understanding (in the log) of what was going on in the consultation.

You cannot pass by default and must show positive evidence of these skills. It is therefore imperative that the consultations on the tape give you the opportunity to demonstrate these skills. A tape which contains only no challenge consultations will therefore not pass the assessment process.

You may record as many consultation sessions as you like in the time available before submitting a tape which you consider demonstrates that you are competent. For this reason you will not be allowed a second attempt within six months of a tape which has failed the process.

If a tape is submitted after the published closing date it will not be assessed and the trainee will not receive a certificate.

### Technical aspects

It is the trainer's responsibility to ensure that the practice has the necessary equipment and expertise to enable the trainee to submit a technically satisfactory tape, i.e. a standard VHS tape with a working camera clock and adequate sound and picture.

### **Consent**

It is the responsibility of the practice to ensure that informed patient consent is obtained for videotaping. A suitable consent form is included in this section.

## Appendix E

### PATIENT'S CONSENT FORM AND INFORMATION SHEET - SUMMATIVE ASSESSMENT

Patient's name .....

#### Consent to Video Recording for Assessment Purposes

We are hoping to make video recordings of some of the consultations between patients and  
 Doctor ..... whom you are seeing today.

The videos are for part of an assessment procedure designed to make sure that all doctors who become GPs are fully competent.

The video is ONLY of you and the doctor talking together. No intimate examination will be done in front of the camera. All video recordings are carried out according to guidelines agreed by the Joint Committee on Postgraduate Training for General Practice, which is the body responsible for the training of GPs.

The video will be seen only by doctors involved in assessment and training of general practice trainees, the tape will then be erased. The tape will be stored in a locked cabinet and is subject to the same degree of confidentiality and security as medical records. The West of Scotland Committee will be responsible for storage of the tape. The tape will be erased as soon as practicable and in any event within 1 year. The results of this work may be used for research and educational purposes.

You do not have to agree to your consultation with the doctor being recorded. If you want the camera turned off, please tell Reception - this is not a problem, and will not affect your consultation in any way.

But if you do not mind your consultation being recorded, we are grateful to you. Improving the assessment of GPs should lead to a better service to patients.

If you wish you may view the tape recording.

If you consent to this consultation being recorded, please sign below. Thank you very much for your help.

Signed            Date .....

Signature(s) of any accompanying person (s) .....

After you have finished seeing the doctor, please sign below to confirm that you are still happy to have the recording used.

Signed            Date .....

*For doctor's use only*

Reference number:

Appendix F

**CONFERENCE OF POSTGRADUATE ADVISERS IN  
GENERAL PRACTICE  
UNIVERSITIES OF THE UNITED KINGDOM**

***SUMMATIVE ASSESSMENT***

***FOR GP REGISTRARS COMPLETING***

***VOCATIONAL TRAINING***

***AFTER 1ST SEPTEMBER 1996***

**October 1995 - 1st issue**

**Index:**

	Page
Why is summative assessment necessary?	2
Can I use the MRCGP ?	2
What are the basic attributes tested by summative assessment?	3
What are the components of summative assessment?	3
What are the outcomes of taking each component of the assessment?	4
What is the role of trainers in summative assessment and how are they being prepared?	4
What do I need to do to prepare myself for taking summative assessment?	5
• MCQ	6
• Structured trainer's report	7
• Written submission of practical work for 1995/96	9
• Assessment of consultation skills (video)	12
Acknowledgments	15

***Important***

- *The term GP registrar is used throughout the document. This title is used for reference to cover doctors who are known under the vocational training regulations as 'general practice trainees'.*
- ***ALL GP REGISTRARS WILL BE ISSUED, VIA THEIR REGION, WITH AN IDENTIFICATION NUMBER WHICH MUST BE USED ON ALL SUBMISSIONS***

## **Introduction**

### **Why is summative assessment necessary?**

There are several reasons why it is now timely to introduce summative assessment at the end of vocational training for general practice. The following are of particular importance:

- To assess the competence of those joining the profession
- To reassure the public and protect patients from doctors whose performance is not adequate
- To reassure individual doctors that they have achieved an agreed minimum standard of competence
- To identify those who are not ready for independent practice and who require further training or need to reconsider their career options

### **Can I use the Membership of the Royal College of General Practitioners (MRCGP)?**

It will be possible and desirable to use the MRCGP examination for exemption from parts of summative assessment.

At present the MRCGP examination MCQ can be used. In future the MRCGP examination will include a consultation skills assessment.

### **What are the basic attributes tested by summative assessment?**

- 1 Adequate knowledge
- 2 Adequate problem solving skills
- 3 Adequate clinical competence
- 4 Adequate consulting skills
- 5 Demonstration of adequate skills in producing a written report of practical work in general practice
- 6 Adequate performance on a wide variety of skills, attitudes and knowledge, confirmed by a trainer's report

### **What are the components of summative assessment?**

There will be four components to summative assessment which will cover the six basic attributes required. They are:

- 1 Multiple Choice Questionnaire (MCQ)  
- (MRCGP examination or UK Regional Advisers (UKRA) assessment)
- 2 Assessments of consultation skills (using video UKRA assessment for 1996 and either MRCGP examination or the UKRA video assessment in the future)
- 3 Written submission of practical work
- 4 Trainer's report

**All 4 components must be passed to complete summative assessment**

### **Note**

*It is intended that all regions will introduce the full package for GP registrars completing vocational training after 01/09/96. A few regions may have to start with only the trainer's report (for GP registrars completing training between 01/09/96 and 01/04/97) and introduce the full package thereafter; this will only occur if regions are unable to secure the resources required. GP registrars will be told by their regional office if this is the position.*

## What are the outcomes of taking each component of the assessment (except MCQ)?

There are 3 possible outcomes

1 **Pass**

2 **Referral (Video or written submission)**

To a 2nd or 3rd level of assessment resulting in either a pass or return of the work to the GP registrar suggesting resubmission (following mentorship by the trainer or course organiser).

**Referral (Trainer report)**

To further consultation with course organisers or advisers to advise on what needs to be done.

3 **Failure**

Ultimate failure is likely to be rare as this is a test of minimum competence. Failure would have serious consequences on a GP registrar's future career. It is anticipated that extended periods of training will be available allowing resubmission. In the rare event of eventual failure, support and counselling could be obtained from course organisers, associate advisers or the regional adviser.

## What is the role of trainers in summative assessment and how are they being prepared?

All trainers will have been briefed regarding all components of summative assessment within their regions. The trainer's primary role is to ensure that the GP registrar is being given the best chance of passing. They will be highlighting any concerns as early as possible in the GP registrar year so that action can be taken. The trainer will enlist the support of others if a GP registrar is not likely to pass all items. GP registrars will then have the support of their trainer, course organiser and associate or regional adviser in reaching the required standard.

## **What do I need to do to prepare myself for taking summative assessment?**

GP registrars should seek guidance from the RCGP (14 Princes Gate, Hyde Park, London SW7 1PU Tel: 0171-581 3232) on taking the MRCGP examination. MRCGP courses are available throughout the country and the College has published a book on the examination (Moore, 1994).

GP registrars should read carefully the following sections on

*MCQ*

*Structured trainer's report*

*Written submission of practical work for 1995/96*

*Assessment of consultations skills (video)*

GP registrars should discuss these at the beginning of their general practice year and if there are any further questions which the trainer is unable to answer, the GP registrar should approach their course organiser.

*Remember this is a test of minimum competence in the wide range of knowledge, skills and attitudes required for an independent practitioner in general medical practice. The vast majority of GP registrars who are conscientiously completing their training in hospital and in their GP registrar year should have no difficulty in passing the assessment.*

**GP REGISTRARS MUST ENSURE THAT THEY KNOW THEIR NUMBER AND THAT IT APPEARS ON ALL SUBMISSIONS.**

All GP registrars must sit the tests in the region in which they are working at the time.

## MCQ

The MCQ paper can be taken as part of the MRCGP examination or within regions as organised by the UKRA. The papers will be of 3 hours duration, testing knowledge and problem-solving skills, and will be held on four separate occasions each year. The specific dates and venues will be published nationally and regionally. The dates set for the UKRA MCQ for 1996 are Wednesday 7th February, Wednesday 1st May, Wednesday 4th September and Wednesday 4th December.

The MCQ paper will consist of 300 items of multiple true or false questions and also extended matching items.

The questions will contain the full range of activities which are part of modern general practice. GP registrars are advised that they can take the assessment having completed 3 months in general practice and when they and their trainer agree they are ready.

Taking the MCQ paper should be straightforward but GP registrars must:

- MAKE SURE THAT THEIR CANDIDATE NUMBER HAS BEEN ENTERED ON THE PAPER
- MARK THE ANSWER SHEETS CLEARLY AND AS DIRECTED (sheets are marked by an optical scanner)

GP registrars who fail the MCQ paper will be permitted to resit on 2 further occasions, if it is part of the MRCGP examination, upon payment of a nominal fee. After this, should a pass not have been achieved, further attempts will incur payment of the full examination fee. GP registrars failing the UKRA MCQ assessment can resit the assessment as appropriate.

Candidates for the MRCGP examination who fail the examination overall, but who satisfy the examiners that their performance in the MCQ paper is adequate, will be issued with a College certificate to this effect. They should present this to the JCPTGP.

## Structured trainer's report

The trainer's report has been developed following a national survey of trainers' views. The standards were produced by a consensus group of experienced trainers. When the standards are reached the GP registrar is considered to be ready for independent practice.

A copy of a blank trainer's report (including guidance for trainers completing the report) will be available from trainers. GP registrars should read this at the beginning of their year and discuss with their trainer when the assessments will be made. Most of the trainer's report will be completed at the beginning of the last 2 months of training (with the exception of specific clinical skills). If a GP registrar changes trainer then the final trainer will seek the advice of the previous trainer(s) in completing the report.

### **The report is divided into 6 sections:**

- 1 patient care (itself divided into general clinical skills, patient management skills and clinical judgement)
- 2 communication skills
- 3 personal and professional growth
- 4 organisational skills
- 5 professional values
- 6 specific clinical skills (this section includes a number of basic diagnostic and therapeutic skills)

The report gives guidance to the trainers on the minimum standards, ie what will constitute a pass/failure. GP registrars can familiarise themselves with these standards by reading the blank report obtained from their trainer.

Trainers are also given guidance as to the best method of assessment under three categories:

- *assessment by observation*
- *assessment by discussion*
- *assessment by specific methods*

All items will need to be completed satisfactorily for the report to be submitted, two months before the completion of training. Trainers have been instructed that whenever there is any doubt about whether or not the GP registrar has reached the necessary standard, repeat observation should be made. Clearly when a trainer is aware that a GP registrar is as yet unable to reach the required standard they will arrange for the appropriate training to occur and ask the advice of a course organiser, associate adviser or regional adviser.

## Written submission of practical work for 1995/96

The written submission of practical work is at present a UKRA assessment exercise and will consist of an audit project. An audit project has been chosen because all general practitioners should be monitoring and improving the quality of care they provide. The ability to carry out an audit is therefore a skill of minimum competence for a GP registrar.

The audit should be submitted typed, in the form stipulated by the proforma on page 10. The audit should be the GP registrar's own work with appropriate support from the practice team. It can be undertaken at any time during the 3 years of vocational training but the work must be relevant to general practice and submitted at least 4 months before the end of vocational training. It would be to the GP registrar's advantage to submit the work as early as possible in case resubmission is necessary.

- Written work should be double spaced, with all pages numbered; the GP registrar candidate number should be on top of each page.
- The submission should be typed in concise English and should normally be no more than 3000 words.

Figures and graphs may be used to support results and conclusions.

The work should be bound or presented in a way that ensures all pages are firmly fixed in order.

- The work will be marked against the schedule on page 11. The schedule will give guidance on what scope of work is appropriate.
- Written submissions will be marked by three assessors independently who will either pass the work or refer the work to two further assessors. If these assessors are not satisfied, the work will be returned to the GP registrar with an indication of where modifications are needed. If problems still occur after resubmission, the work will be referred to 2 assessors, external to the region.
- Trainers or course organisers will give advice on the type of submission and the scope of work which is appropriate.
- GP registrars should discuss with their trainers or course organiser the resources they will require to carry out the audit, including protected time.

## Proforma

Audit Title:

Candidate Number:

Please ensure that your candidate number has been entered on the title page and on each subsequent page

Trainer name:

Practice address:

Practice list size (approx):

The enclosed audit project is the work of the GP registrar named above and was carried out during the general practice component of training.

Trainer signature:

- What is the title of your audit project?
- Why did you choose it?
- What criteria have you chosen?
- Why did you choose them?
- What preparation and planning did you undertake for your audit project?
- Summarise relevant data.
- What conclusion do you draw from these?
- Detail the changes you propose as a result of your audit project.

## MARKING SCHEDULE

QUESTION	CRITERION	CRITERIA PRESENT
<u>Why was the audit done?</u>	<u>Reason for Choice</u>  Should be clearly defined and reflected in the title. Should include potential for change.	
<u>How was the audit done?</u>	<u>Criteria Chosen</u>  Should be relevant to the subject of the audit. Should be justified eg literature.  <u>Preparation and Planning</u>  Should show appropriate teamwork and methodology in carrying out audit. If standards are set they should be appropriate and justified.	
<u>What was found?</u>	<u>Interpretation of Data</u>  Should use relevant data to allow appropriate conclusions to be drawn.	
<u>What next?</u>	<u>Detailed Proposals for Change</u>  Should show explicit details of proposed changes.	

*A satisfactory GP registrar audit report must include all 5 criteria to pass*

## Assessment of Consultation Skills (Video)

This is currently available as devised by the UKRA and will also be available separately as part of the MRCGP examination in the future.

- This assessment will be based on video recordings of patients and can be submitted after 3 months in the general practice year but GP registrars are advised that it may be wise to delay until after 6 months. The trainer will advise the GP registrar on the correct time for him/her to submit their tape.
- Regions will publish dates on which they will accept tapes. All GP registrars should ensure that they have submitted their tape on a date that is at least 3 months before the end of the GP registrar year.
- Each GP registrar should submit to their regional office a 2 hour tape for the UKRA assessment.

### **Tapes are to be submitted in standard VHS format.**

If the GP registrar or trainer has a problem with the format of the tape, they are advised to discuss this early with either the regional adviser or the associate adviser. Blank tapes will be supplied by the regional office.

- All consultations should be submitted except where the patient withholds consent.
- Verification of the identity of the doctor on the tape will be required. You will be advised from your regional office how this should be done.
- Tapes will be accepted after January 1996

## **PARTICULAR ATTENTION SHOULD BE PAID TO THE FOLLOWING POINTS**

- 1 **OBTAINING FULL CONSENT FROM THE PATIENT.** The trainer will have the standard form and instructions for this.
- 2 **SENSITIVE CLINICAL EXAMINATIONS SHOULD NOT BE IN THE PICTURE,** but the tape must be kept running so the sound is recorded.
- 3 **GOOD QUALITY PICTURES AND SOUND.** The trainer or course organiser should be able to help the GP registrar with this. **TAPES OF POOR QUALITY WILL BE RETURNED FOR RESUBMISSION**
- 4 **INCOMPLETE CONSULTATIONS SHOULD BE DELETED**
- 5 **KEEP THE CAMERA CLOCK ON TO AID THE ASSESSOR**
- 6 **COMPLETE A LOG ENTRY FOR EACH CONSULTATION.**  
Logbooks can be obtained via the GP registrar's trainer, day release course or regional office. GP registrars will be reviewing and commenting on their own consultations in this logbook. This will also give them an opportunity to check quality.
- 7 **COMPLETED TAPES SHOULD BE SENT VIA REGISTERED POST OR BY HAND (each regional office will issue local instructions) TO THE REGIONAL OFFICE FOR REGISTRATION, STORAGE, PROTECTED DISTRIBUTION FOR ASSESSMENT AND CLEARING AFTER ASSESSMENT.**

- All tapes will be reviewed by 2 assessors and, if necessary, referred to further levels of assessment. It is anticipated that only a small proportion of GP registrars will be referred for further assessment. If doubts occur after the second level of assessment, the tape will be viewed by 2 assessors external to the region. The criteria will be as follows:
  - **Listening.** The GP registrar should identify and elucidate the reason or reasons for the patient's attendance. Credible and mutually acceptable plans should be negotiated with the patient.
  - **Action.** The GP registrar should take appropriate action to identify the patient's problem or problems. Investigations and referral should be reasonable. Help should be sought when necessary. The patient's problem should be managed appropriately.
  - **Understanding.** The GP registrar should demonstrate in the logbook that he or she understands the process and outcome of the consultation. Individual action should be explained. Obvious shortcomings in the consultations should be identified and relevant background should be mentioned.
  - **Error.** If a major error is noted in patient management, or if a series of minor errors is noted, the GP registrar may be referred. (A major error is one which causes actual or potential harm to the patient; a minor error causes inconvenience only.)

#### Further Information

- The JCPTGP will be issuing administrative instructions on the certification process incorporating summative assessment.
- The details of the assessment may be subject to minor change or variation. If a GP registrar has any doubts, they should seek advice from the regional adviser's office.
- GP registrars should remember the importance of presenting original data honestly acquired. Fabricated data may lead to disciplinary procedures via the GMC. Extracts from papers etc should have acknowledgement of source.

October 1995

Prepared by Dr D B Percy  
Regional Adviser in General Practice and Associate Dean  
South and West Regional Health Authority (Wessex)  
on behalf of the UKRA Summative Assessment Working Group

## ***Acknowledgments:***

Production of this information pack relied heavily on information provided by

- The Summative Assessment Working Party of the UK Advisers under the chairmanship of Dr John Hasler, OBE, MD, FRCGP.
- Prof Stuart Murray, MD, PhD, FRCGP, FRCP(G), FRCP(E), Dr Malcolm Campbell, FRCGP, and Dr Murray Lough, MRCGP of West of Scotland Postgraduate Medical Education Board.
- Dr Neil Johnson, MA, MSc, MRCGP, MRCP, Research Fellow, University of Oxford Department of Public Health and Primary Care.
- The Chairman and Officers of the Joint Committee for Postgraduate Training for General Practice.
- Examination Department of the Royal College of General Practitioners.

Ref: 'The MRCGP Examination and Summative Assessment' published by the Royal College of General Practitioners - May 1995.

**REFERENCES**

- Arborclius E, Krakau I, Bremberg S. 1992. *Key factors in health counselling in the consultation.* Journal of Family Practice. 9: p488-493.
- Bain J E, Mackay N S D. 1993. *Videotaping general practice consultations {letter}.* British Medical Journal. 307: p504-5.
- Bain J E, Mackay N S D. 1995. *Videotaping of general practice consultations. {letter}.* British Medical Journal. 311: p952.
- Bain J E, Mackay N S D. 1995. *Patients' assessment of trainee general practitioners.* Journal of Medical Education. 29: p91-96.
- Baird A G, Gillies J C M. 1993. *Videotape assessment is threatening {letter}.* British Medical Journal. 307: p60.
- Baker R. 1990. *Development of a questionnaire to assess patient satisfaction with consultation in general practice.* British Journal of General Practice. 40: p487-490.

Banks I. 1994. *Speech to English National Trainee Conference.*

Bass M J, Buck C, Turner L. 1986. *The physician's actions and the outcome of illness in family practice.* Journal of Family Practice. 23: p43-47.

Benner D. 1984. *From novice to expert: excellence and power in clinical nursing practice.* Addison-Wesley, Menlo Park, USA.

Berliner D C. 1987. *Ways of thinking about students and classrooms by more and less experienced teachers.* In: *Exploring teachers thinking.* Cassell. London.

Boardman et al. 1987. *(letter).* Journal of Royal College of General Practitioners. 37: p134.

Bulstrode C, Holsgrove G. 1996. *Education for educating surgeons.* British Medical Journal. 312: p326-327.

Calman K C (Chairman). Chief Medical Officer. 1995. *Review of guidance on doctors' performance.*

Campbell L M, Murray T S. 1990. *Trainee assessment - a regional survey.* British Journal of General Practice. 40 (341): p507-509.

Campbell L M, Howie J G R, Murray T S. 1993. *Summative assessment: A pilot project in the West of Scotland.* British Journal of General Practice. 43: p430-434.

Campbell L M, Murray T S. 1994. *Leicester Assessment Package {letter}.* British Journal of General Practice. 44 (388): p535.

Campbell L M, Murray T S. 1995a. *Dealing with Duffers {letter}.* Lancet 346: p1164.

Campbell L M, Sullivan F, Murray T S. 1995b. *Videotaping of general practice consultations: effect on patient satisfaction.* British Medical Journal. 311: p236.

Campbell L M, Howie J G R, Murray T S. 1995c. *The use of videotaped consultations in the assessment of general practice trainees.* British Journal of General Practice. 43: p430-434.

Campbell L M, Murray T S. 1995d. *Reliability of the MRCGP viva. (letter).* British Medical Journal. 311: 1572-3

Campbell L M, Murray T S. 1996. *Summative assessment in the West of Scotland - the results of the first 3 years.* British Journal of General Practice. 46: 411-4

Carney P A, Dietrich A J, Freeman D H Jr, Mott L A. 1995. *A standardised-patient assessment of a continuing medical education program to improve physicians' cancer-control clinical skills.* Academic Medicine. 70 (1): p52-58.

Carney T. 1992. *A national standard for entry into general practice [editorial].* British Medical Journal. 305: p1449-1450.

Colliver J A, Vu N V, Markwell S J, Verhulst S J. 1991. *Reliability and efficiency of components of clinical competence assessed with five performance-based examinations using standardised patients*. Medical Education. 25: p303-310.

Comstock L M, Hooper E M, Goodwin J M, Goodwin J S. 1982. *Physician behaviours that correlate with patient satisfaction*. Journal of Medical Education. 57: p105-112.

Copolla W, Chairman of Trainee subcommittee, General Medical Services Committee. 1995. *English National Trainee Conference (Unpublished Speech)*.

Cox J, Mullholland H. 1993. *An instrument for assessment of videotapes of general practitioners' performance*. British Medical Journal. 306: p1043-1046.

Cox K. 1990. *No Oscar for OSCA*. Medical Education. 24 (6): p540-545.

De Gruijter D N M. 1985. *Compromise models for establishing examination standards*. *Journal of Educational Measurement*. 22: p263-266.

Department of Health. 1989. *General Practice in the National Health Service. The 1990 Contract*. Health Departments of Great Britain.

Dowaliby F J, Andrew B J. 1976. *Relationships between clinical competence ratings and examination performance*. *Journal of Medical Education*, 51 (3): p181-188

Dwyer J W, Derweiller N L, Kosch S G. 1988. *Medical students and comprehensive patient care: Attitudes, perceived competence and demonstrated ability*. *Medical Education*. 22: p19-26.

Edelstein D R, Ruder H J. 1990. *Assessment of clinical skills using videotapes of the complete medical interview and physical examination*. *Medical Teacher*. 12 (2): p155-162.

Elstein A S, Shulman L S, Sprafka S A. 1978. *Medical problem solving*. Harvard University Press, Cambridge, Massachusetts.

Ferrell B G. 1995. *Clinical performance assessment using standardised patients*. *Journal of Family Medicine*. 27: p14-19.

Fraser R C, McKinley R K, Mullholland H. 1992. *Assessment of consultation competence in general practice: the Leicester assessment package*. In: Harden R M, Hart I R, Mullholland H, eds. *Approaches to the assessment of clinical competence*. Dundee: Centre for Medical Education, 1992: p192-198.

Fraser R C, Mckinley R K, Mulholland H. 1994. *Consultation competence in general practice: establishing the face validity of prioritised criteria in the Leicester assessment package*. *British Journal of General Practice*. 44: p109-113.

Freeman J, Byrne P S. 1976. *The assessment of postgraduate training in general practice*. Guildford: Society for Research into Higher Education, University of Surrey.

General Medical Council. 1980. Education Committee: *Recommendations on basic medical education*. GMC London.

General Medical Council. 1994. *Guidance for doctors on the use of videorecordings of consultations between doctors and patients, and of other medical procedures for the purposes of training and assessment*. London: GMC.

General Medical Council. 1993. *Tomorrow's Doctors: Recommendations on Undergraduate Medical Education*. GMC London.

General Medical Services Committee. 1992. *Your choices for the future*, page 101. British Medical Association.

Godlee F. 1991. *RCGP - Examining the exam*. British Medical Journal. 303: p325-328.

Gorman T P, White J, Orchard A. *Language performance in schools: secondary survey report No 2*. London HMSO.

Grol R, de Maeseneer J, Whitfield M, Mookink H. 1990. *Disease centred versus patient-centred attitudes: comparison of general practitioners in Belgium, Britain and the Netherlands*. *Journal of Family Practice*. 7: p100-103.

Hannay D R. 1979. *Postgraduate examinations for general practice in Canada and the United Kingdom*. *Lancet*, Vol 2 (8132): p28-30.

Harden R M, Gleeson F A. 1979. *Assessment of medical competence using an objective structured clinical examination*. *Medical Education*. 13: p39-54.

Hasler J. 1994. *Chairman's statement, UK Conference of Regional Advisers, Cambridge*.

Havelock P, Hasler J, Flew R, McIntyre D, Schofield T, Toby J. 1995. *Professional Education for General Practice*. Oxford University Press.

Havelock P, 1994. *Personal communication*

Hays R B, Jones B F, Adkins P B, McKain P J. 1990a. *Analysis of videotaped consultations to certify competence.* Medical Journal of Australia. 152 (11): p609-11.

Hays R B. 1990b. *Assessment of general practice consultations: content validity of a rating scale.* Journal of Medical Education. 24: p110-116.

Hays R B, van der Leuten C, Fabb W E, Spike N A. 1995. *Longitudinal reliability of the Royal Australian College of General Practitioners Certification Examination.* Medical Education. 29: p317-321.

Henbest R J, Stewart M. 1990. *Patient centredness in the consultation. 2. Does it really make a difference?* Journal of Family Practice. 7: p28-33.

Herbers J E et al. 1989. *How accurate are faculty evaluations of clinical competence?* Journal of General Internal Medicine. 4 (3): p202-208.

Hojat et al. 1993. *Is the glass half full or half empty? A re-examination of the associations between assessment measures during medical school and clinical competence after graduation.* Academic Medicine (2 Suppl): pS69-76.

Houston H L, Reilly T A, Sibert J R, Davies D P. 1995. *Child health surveillance: development of a multicentre course and its evaluation*. Medical Education. 29 (3): p242-246.

Hubbard et al. 1965. *An objective evaluation of clinical competence*. New England Journal of Medicine. 272: p1321-1337.

Irvine D H, Gray D J P, Bogle I G. 1990. *Vocational training for general practice: the meaning of 'satisfactory completion' {letter}*. British Journal of General Practice. 40: p434.

Irvine D and Irvine S (Editors). 1991. *Making sense of Audit*. Radcliffe Medical Press.

Joint Committee on Postgraduate Training for General Practice. 1987. *Assessment Working Party Report*. London: JCPTGP.

Joint Committee on Postgraduate Training for General Practice. 1992. *Report of the Working Party on Summative Assessment*. London: JCPTGP.

Joint Committee on Postgraduate Training for General Practice. 1995. *Summative assessment of vocational training in general practice. A second position paper.* London: JCPTGP.

Jones T V, Gerrity M S, Earp J. 1990. *Written case simulations: do they predict physicians' behaviour.* *Journal of Clinical Epidemiology.* 43 (8): p805-815.

Kelly D R, Murray T S. 1991. *Twenty years of vocational training in the West of Scotland: The Practice component.* *British Journal of General Practice.* 41 (353): p492-495

Kelly M H. *Effectiveness of continuing medical education.* Submitted for publication.

Kemple T. 1995. *Your recertified future - part 5.* *British Journal of General Practice.* 45: V11.

Livingstone S A, Zieky M J. 1982. *Passing Scores*. Education Testing Service, Princeton, New Jersey.

Lough J R M, McKay J, Murray T S. 1995. *Audit and summative assessment: a criterion-referenced approach*. British Journal of General Practice. 45: p607-609.

Mackay H A F. 1987. *Videotaped consultations.{letter}*. Journal of Royal College of General Practitioners. 37: p181.

Maguire P. 1985. *Psychological barriers to care of the dying*. British Medical Journal. 291: p1711-1713.

Maguire P. 1989. *Assessing clinical competence [editorial]*. British Medical Journal. 298: p4-5.

Martin E, Martin P M L. 1984. *Patients views on video recording of consultations*. British Journal of General Practice. 34: p607-610.

Millar T, Goldberg D P. 1991. *Link between the ability to detect and manage emotional disorders : a study of general practitioner trainees.* British Journal of General Practice. September, Vol. 41 (350): p357-359.

Moorhead R, Winfield H. 1990. *Teaching counselling skills to fourth year medical students: a dilemma concerning goals.* Family Practice. 8: p343-346.

Mullholland H, Tombleson P. 1990. *Assessment of the general practitioner.* British Journal of General Practice. Vol: 40 (335): p252-254.

Neighbour R. 1993. *The inner consultation.* Oxford University Press.

Newble D I, Swanson D B. 1988. *Psychometric characteristics of the objective structured clinical examination.* Medical Education. 22 (4): p325-334.

Norcini et al. 1985. *Reliability, validity and efficiency of multiple choice question and patient management problem item formats in assessment of clinical competence.* Medical Education. 19 (3): p238-247.

Norman G R, Neufeld V R, Walsh A, Woodward C A, McConvey G A. 1985. *Measuring physicians' performance by using simulated patients*. Journal of Medical Education. 60: p925-934.

Pendleton D, Schofield T, Tate P, Havelock P. 1984. *The consultation*. Oxford University Press.

Pieters H M, Touw-Otten F W, De Melker R A. 1994. *Simulated patients in assessing consultation skills in general practice vocational training: a validity study*. Medical Education. 28 (3): p226-233.

Pringle M, Robins S, Brown G. 1984. *Assessing the consultation: methods of observing trainees in general practice*. British Medical Journal. 288: p1659-1660.

Pringle M, Stewart-Evans C. 1990. *Does awareness of being videorecorded affect doctors' consultation behaviour?* British Journal of General Practice. 40: p455-458.

Rabinowitz H K. 1987. *The modified essay question: an evaluation of its use in a family medicine clerkship.* Medical Education. Vol. 21 (2): p114-118.

Rethans J J, van Leeuwen Y, Drop R, van der Vleuten C, Sturmans F. 1990. *Competence and performance: two different concepts in the assessment of quality of medical care.* Journal of Family Practice. 7 (3): p168-174.

Rethans J J E, Van Boven C P A. 1987. *Simulated patients in general practice: a different look at the consultation.* British Medical Journal. 294: p809-812.

Rethans J J, Sturmans F, Drop R, van der Vleuten C. 1991. *Assessment of the performance of general practitioners by the use of standardised (simulated) patients.* British Journal of General Practice. 41 (344): p97-99.

Roberts J, Norman G. 1990. *Reliability and learning from the objective structured clinical examination.* Medical Education. 24 (3): p219-223.

Royal College of General Practitioners. 1988. *Rating Scales for the assessment of Vocational trainees*. Occasional Paper No. 40.

Savage R, Armstrong D. 1990. *Effect of a general practitioner's consulting style on patients' satisfaction: a controlled study*. *British Medical Journal*. 310: p968-970.

Servant J B, Mathieson J A B. 1986. *Video recording in general practice: the patients do mind*. *Journal of Royal College of General Practitioners*. 36: p555-556.

Schon, D. A. 1983. *The reflective practitioner*. Basic Books, New York.

Scully G. 1995. *Dealing with Duffers*. *Lancet*. 346: p720.

Southgate L. 1993. *Guidelines on the use of videotaped consultations*. London, Royal College of General Practitioners.

Taylor J, Taylor D. 1988. *The assessment of vocational training in general medical practice*. York: University of York, Centre for Health Economics.

Tamblyn R M, Klass D J, Schnabl G K, Kopelow M L. 1991. *The accuracy of standardised patient presentation*. Medical Education. 25: p100-109.

Tylee A et al. 1987. *Videotaped consultations {letter}*. Journal of Royal College of General Practitioners. 37: p320.

Wakeford R, Southgate L, Waas V. 1995. *Improving oral examinations: selecting training and monitoring examiners for the MRCGP*. British Medical Journal. 311: p931-935.

Walker R, Walker B. 1987. *Use of the objective structured clinical examination for assessment of vocational trainees in general practice*. Journal of Royal College of General Practitioners. 37: p123-124.

Williams et al. 1987. *Direct, standardised assessment of clinical competence*. Medical Education. 21 (6): p482-489.

Vu N V, Marcy M M, Colliver J A, Verhulst S J, Travis T A, Barrows H S.  
1992. *Standardised (simulated) patients accuracy in recording clinical  
performance check-list items.* Journal of Medical Education. 26: p99-104.

