PRACTICE OBSERVED

Practice Research

Entry to general practice training

JAMES ROBERTS, JAMES FREEMAN

Abstract

Since 1979 general practitioner trainees in the North West Region of England have been assessed on knowledge and ability by several tests, including the multiple choice test. Trainers in the region require a score in the MCQ of 45% before taking a trainee into the practice. More British graduates have achieved this score than overseas graduates. The difference was statistically significant. Only applicants who show adequate factual recall should be appointed as trainees.

Introduction

The purpose of mandatory training for general practice is to raise the level of practice in Britain. The training is laid down in terms of experience offered by hospital and general practice trainers and not in terms of benefit gained by the trainee. Once a trainee has satisfactorily completed training a certificate is granted which enables the trainee to enter general practice in the National Health Service, even though he may be grossly incompetent.

For some years trainers in the north west of England have realised that some applicants for trainee posts do not have a standard of factual recall of clinical matters sufficiently high for them to benefit from the trainee year. (In any educational exercise at the advanced or higher levels participants must have attained a taxonomic level, basic to that exercise, to participate effectively and to benefit. This is the logic underlying the concept of all minimum course requirements implemented in most institutions of higher education.) The trainers were also aware that the interview and taking up references rarely brought out such basic deficiencies, to the possible detriment of patients in the practice when the applicant was appointed.

Trainees in the day release courses have expressed concern that visiting speakers are occasionally asked questions by members of their groups on matters that should be assumed to be undergraduate common knowledge and that such questions gave vocational training and trainees a poor image. Furthermore, group work was seriously inhibited by the lack of a common basis of clinical factual recall from which intelligent discussion could proceed. Their time was being wasted and a feeling of hopelessness was engendered at the low degree of achievement in these sessions. Since September 1979, therefore, the trainers in the region have agreed that they will not accept as trainees applicants who are not in vocational schemes who cannot show that they have reached a specific level of adequate recall of clinical facts. Trainees in vocational schemes are selected out by other means and earlier in their training. They normally voluntarily undergo the assessment with universally good results.

A system has been set up whereby the trainee applicant, as part of a pre-employment interview, completes a multiple choice test that covers the specialties experienced in general practice. The proportional content of the paper was determined by lists constructed from the General Register Office and Royal College of General Practitioners Morbidity Surveys. The test has 220 questions—36 psychiatry, 40 paediatrics and infectious diseases, 40 medical, 24 therapeutics, 24 obstetrics and gynaecology, 16 ear, nose, and throat, and ophthalmics and dermatology, 20 social medicine, and 20 surgery.[1] It is not negatively marked and is the type in which the candidate has to select a single correct answer from a group of five choices.[2] The paper is personally administered and supervised, explanation and reassurance are therefore at hand, and the applicant is allowed whatever time is necessary to answer the questions. The fact that an error percentage is not deducted removes the pressure from the examination. The choice of one in five answers gives enough flexibility to permit this approach and, as the results show, still obtain a good spread of scores and discrimination at the lower end and middle of the scale. Initially, the trainers were advised that the applicant was acceptable if he or she scored 35%, but course organisers and trainers indicated that this was too low and in November 1982 the minimum required score was raised to 45%.

Whether or not the applicant achieves the required score he or she is given a summary of his own results in each of the specialties and those of groups of trainers, British graduates, and overseas graduates. If the trainee applicant is successful the trainer also receives a copy. Apparent deficiencies discovered by the multiple choice test in the various medical specialties incorporated in the test may then be carefully assessed and remedied where necessary. The unsuccessful applicant is given the opportunity to revise over six months, in his own time, and to retake the assessment. A low result does not permanently exclude a candidate.

Department of General Practice, University of Manchester, Gateway House, Piccadilly South, Manchester M60 7LP
JAMES ROBERTS, MB, FRCGP, regional adviser in general practice North West Region
JAMES FREEMAN, PHD, DIPSCON, senior research officer

Results and discussion

Of 602 candidates who undertook the assessment, 110 scored under 45%. At the other end of the scale there were 75 who scored over 65%. It was clear from the outset that there was a definite distinction between British graduates and overseas graduates (table I). One candidate scored under 20%.

TABLE I—*Scores for British and for overseas graduates on multiple choice test*

	No who scored under 35%	No who scored under 45% (includes previous column)	No who scored over 65%
British graduates (n = 257)	Nil	2	69
Overseas graduates (n = 345)	30	108	6

The difference between the two groups is statistically significant by the χ^2 test: $p < 0{\cdot}0001$.

The term "British graduate" includes only graduates of the universities of Great Britain, excluding Eire, irrespective of country of origin of the graduates. Undoubtedly trainers and course organisers are much happier with the standard of trainee, although course organisers would like to see the required score of 45% raised.

Although the multiple choice test does not itself establish that the doctor is incapable of benefiting from the trainee year, other evidence showed that almost invariably trainees who did badly on factual recall also did badly in other areas important in general practice.[3] The multiple choice test was administered as part of a range of tests including the multiple essay questions to measure problem solving skills in patient management, the traditional essay questions to assess management skills and similar high level taxonomic abilities, personality measures, culture—fair measures of problem solving ability—and measures of convergence and divergence in several modalities. There were high correlations among these measures, particularly at the lower end of the scale. The multiple choice test tended to have the highest score.[4]

Follow up studies of trainee entrants carried out over seven years showed consistent findings. The test battery including the multiple choice and multiple essay questions was administered before and after the training course. Table II shows the results obtained when two groups were compared. Group A was composed of trainee entrants whose pre-entry multiple choice test scores were 45 or below. Group B, matched with group A for age and experience after registration, was composed of trainee entrants whose pre-entry multiple choice test scores were 50 or above. These groupings were made irrespective of university attended and country of origin. The consistent differences in the scores of these groups after training is striking. Group B obtained a mean score of 63·5 after the course—statistically significantly greater than the mean of 39·0 obtained by group A. The mean increase of 7·5 for group B is statistically significantly greater than the 3·5 for group A. These findings are partly explained by the fact that the goals of vocational training are focused on problem solving skills rather than on increasing recall of clinical facts. Interestingly the scores for the multiple essay question for both groups showed similar patterns to those observed on the multiple choice test. Group B showed a statistically significantly greater increase in scores for the essay question after the course.

TABLE II—*Comparison of low scorers with high scorers on the multiple choice test (MCQ) before entering the training course*

	Mean (range) entry MCQ score	Mean (range) MCQ after the course	Mean increase in score between entry and after course
Group A (n = 100) MCQ score = 45 or below	35·5 (31·0-45·5)	39·0 (34·0-52·0)	3·5
Group B (n = 100) MCQ score = 50 or above	56·0 (52·0-69·5)	63·5 (54·5-71·5)	7·5

The difference between the entry scores is statistically significant in favour of group B by the Mann Whitney U test: $p < 0{\cdot}01$.
The difference between the scores after the course is statistically significant in favour of group B by the Mann Whitney U test: $p < 0{\cdot}001$.
The difference between the increase in score between entry and after the course between the two groups is statistically significant in favour of group B by the Mann Whitney U test: $p < 0{\cdot}01$.

The multiple choice questions were chosen to reflect the range and the appropriate depth of knowledge for general practice. The items for the multiple essay question were specifically on problems of management of patients in general practice. The test items in both measures were updated annually when necessary to reflect, for instance, changes in tests and treatments. Care was taken, however, to ensure that the appropriate level of difficulty and internal consistency of the tests were maintained.

A criticism of measures of this type is that they are academically orientated and do not necessarily reflect the professional performance of the doctor. In this study this criticism was met because during postgraduate training the tutors were asked to rate each trainee on nine criteria—this was essentially the rating scale adapted from the technique developed by McGuire.[4] The criteria used on the rating scale were: (*a*) information gathering; (*b*) problem solving; (*c*) clinical judgment; (*d*) relationship to patients; (*e*) continuing responsibility; (*f*) emergency care; (*g*) relationship with colleagues; (*h*) professional values; (*i*) overall competence. A 12 point scale was used for each criterion, where ratings of 1-3 are poor, 4-6 marginal, 7-9 good, and 10-12 excellent.

Table III clearly shows the statistically significant differences between group A and group B on the tutors' assessments. Indeed, it was the poor performance during the course of some trainees in this group that motivated making these assessments.

TABLE III—*Comparison of tutors' course assessments of low scorers and high scorers on multiple choice test (MCQ) before entering course*

	Mean overall rating (Manchester scales)	Classification (McGuire*)	Standard deviation	Range
Group A (n = 100) MCQ score = 45 or below	4·50	Poor/marginal	3·45	2-9
Group B (n = 100) MCQ score = 50 or above	10·25	Excellent	1·12	8-12

The difference between the two groups in tutors' ratings on the evaluation scales is significantly different in favour of group B by the Mann Whitney U test: $p < 0{\cdot}001$.

At the end of training the trainees from both groups completed a questionnaire giving details of their teaching-learning experience during vocational training. It is fascinating that the most stringent criticisms and evaluation came from the members of group B, who criticised the adequacy of the teaching methods, practice facilities (rooms and equipment), and quality of record keeping both for patient care and for teaching. Little criticism came from group A, the low scorers on the multiple choice test before entry. Attendance at further education and training lectures and seminars was reported more often from group B.

Although this may be seen as a function of the motivation of this group rather than a result of scores on multiple choice tests, the multiple choice test score is a reasonably accurate predictor. Fifty eight members of group B reported that they intended to take the membership examination of the Royal College of General Practitioners against five from group A, and some 50 members of group B were successful against none from group A.

Table II shows a wide range of scores after the course in both groups. This is clearly a function of the influence of the tutor on the trainee—a factor which we have examined.[4] This finding, which has consistently affected results in this topic, however, in no way obscures the present study.

Reasons often offered for poor performance on the multiple choice test are: (*a*) not understanding the question owing to the language problem (but this is the language trainees will be working in, interviewing patients, asking questions, making diagnoses, and managing and discussing treatment); (*b*) lack of familiarity with the technique of the multiple choice test: the trainees were given several examples of how the questions should be tackled just before the examination. A strict time limit was not imposed so as to allow the candidate to settle down without becoming unduly nervous. The candidates were informed that errors were not being penalised in a negative scoring system. The examiner ensured that the candidates understood the questions by going through several examples thoroughly with them at the beginning of the examination.

Research has shown that if extra time is allowed the overall score improves. The success of this is shown by the fact that scores for recall of clinical facts were higher in the multiple choice test than in the traditional essay question—a technique with which some candidates claimed they were more familiar. Thus familiarity with the technique was not reflected in the percentage scores. Overall performance on the multiple choice test in terms of relative scores was invariably better than performance on the multiple essay question, the traditional essay question, and other measures.

The true reasons are probably poor undergraduate training or lack of recent experience of general practice specialties, and, so, loss of distant recall. This applies to doctors who have been many years in a specialty—for example, anaesthetics—and who have to change course. In the past these doctors have been considered not good enough for the specialty but "good enough for general practice." This is no longer so, without further revision and training before the trainee year. To safeguard the future of general practice all applicants for trainee posts in general practice should be required to show themselves capable of benefiting from the trainee year before entering on that year.

References

1 Freeman J, Roberts J, Metcalfe D, Hillier V. *The influence of trainers on trainees.* Occasional paper 21. London: Journal of the Royal College of General Practitioners, 1982.
2 Anderson J. *The multiple choice question in medicine.* London: Pitman, 1982.
3 Freeman J, Roberts J. *Retrospect and prospect: a study of vocational training.* Report to the DHSS, 1979.
4 Freeman J, Byrne PS. *The assessment of vocational training.* Monograph 20. Guildford: University of Surrey, SRHE, 1975.
5 Siegel S. *Non-parametric statistics.* New York: McGraw-Hill, 1956.
6 McGuire CH. *An evaluation model for professional education.* Chicago College of Medicine, University of Illinois, 1967.

(*Accepted 24 October 1983*)

Do trainees see patients with chronic illness?

JOHN C HASLER

Abstract

From 1976 to 1979 trainees in the Oxford region had their clinical work monitored while in their training practices in relation to 16 conditions requiring long term supervision. Although the mean numbers of patients seen were not greatly different from national figures, the range identified some trainees whose number of patients was very low. Consultation rates and follow up figures indicated that many patients did not return to the trainees again.

Medical School Offices, John Radcliffe Hospital, Headington, Oxford OX3 9DU
JOHN C HASLER, MD, FRCGP, regional postgraduate adviser and clinical lecturer in general practice

Introduction

Vocational training for general practice has been in operation for well over a decade, and in 1982 a three year postregistration period of training became mandatory for those doctors wishing to practise as principals in the National Health Service. Although there have been many papers on the educational aspects of training, comparatively little has been published on the clinical work that trainees actually do in their training practices.

The year in the training practice consists of a combination of service work interspersed with educational activities: for most, if not all, trainees service work occupies most of the time. In a recent national survey carried out for the Exeter National Trainee Conference 806 out of 1335 trainers were providing less than two hours' average teaching time a week in the opinion of their trainees.[1] On top of this most trainees attend a course for at least one half day a week for a minimum of 30 weeks in the year.

We already have evidence in the Oxford region that trainers have difficulty in choosing what to teach (J Pearce, unpublished observations).[2] Much teaching tends to be based on cases that

the trainee has seen, and therefore it is important to know something about the pattern of work.

It cannot be assumed that trainees' clinical experience is adequate, nor is it necessarily the same as their trainers'. Twelve months is a relatively short period, and the trainee does not automatically see and manage the patients who present to his trainer and the other partners. Patients with long term problems often want to see their own doctor and if he is away will wait for him to return. By comparison, there have been suggestions that trainees see proportionately more acute minor illness than their trainers.

There have been accounts of trainees' clinical work either from individuals[3 4] or from a small group[5 6] published in the past 14 years. In only three of them is there enough detail to be able to draw conclusions in any detail. An editorial in the *Journal of the Royal College of General Practitioners* in 1979 pointed out the crucial importance of being clear about the nature of clinical experience in the general practice training year.[7]

It has been suggested that one area of relative deficiency might be chronic problems and conditions needing long term supervision. This impression is confirmed by three of the papers already mentioned.[4 5 6] Accordingly, I decided to investigate this in the Oxford region's vocational training schemes.

Study

Seventy five of 109 trainees entering training practices in the region from October 1976 to August 1978 were selected to take part. These 75 were not chosen according to any plan. Four trainees did not wish to take part, a further 11 dropped out during the study, and data from one other were excluded because he had not understood the instructions. This left 59 trainees, whose characteristics were similar to those of the total 109 trainees entering training in the region, except that no trainees in the study graduated in the Indian subcontinent, compared with 9·6% of the trainees entering the region during the study period. This figure in turn was roughly half the percentage for trainees in England and Wales as a whole in 1979. This fact apart, the trainees in the study were similar in relation to age, sex, and duration of previous hospital experience to trainees in England as a whole: two thirds were on three year rotations and the remainder on self constructed programmes. Most attachments were for 12 months but some were for less. Two trainees did two periods in different practices and their periods were analysed twice.

Results

The trainees identified all patients who consulted them for one of the conditions shown in table I. Patients with one of these conditions, but consulting for something else, were excluded. This was because it was important to be clear that the trainee actually participated in the management of the condition. It was also discovered in the pilot study that the trainee could not always easily identify from the records patients who suffered from one of the conditions unless they presented it at the consultation. In most cases the diagnosis was not in doubt, but the trainees were given verbal and written guidance. For each patient, the trainee completed a recording form which remained in the clinical record until the end of the trainee's attachment, and each time the trainee saw the patient again, for whatever reason, an entry was made of the presenting problem at that consultation.

To check on the efficiency (accuracy) of the trainees' recruiting, batches of their prescriptions were obtained from the Prescription Pricing Authority with their permission, and patients on certain drugs were cross checked at the end against patients they had recruited. This enabled an efficiency (or accuracy) rating to be established for each trainee. Data relating to numbers of patients were excluded from trainees who achieved an efficiency rating of 70%, or less. Data relating to follow up, however, were included from all relevant trainees as once a patient had been recruited the research associate collecting the recording forms at the end checked that all subsequent consultations by the trainee had been entered.

Table II shows the numbers of patients with eight of the conditions recruited by the trainees achieving a recording efficiency of over 70%.

TABLE I—*Conditions requiring long term supervision used in the study*

Angina	Mental handicap
Asthma	Oral contraceptive prescription
Cardiac failure	Osteoarthritis
Depression	Physical handicap
Diabetes	Pregnancy
Epilepsy	Psoriasis
Hypertension	Rheumatoid arthritis
Hypothyroidism	Recurrent urinary tract infection

TABLE II—*Numbers of patients recruited by 36 trainees in first or only six months and by 21 trainees in a further period of five or six months*

Condition	No of patients							
	First or only six months (n = 36)				Further period five or six months (n = 21)			
	Mean	0	1-4	5+	Mean	0	1-4	5+
Asthma	12·7	0	6	30	7·9	0	3	18
Depression	13·7	0	4	32	9·3	0	5	16
Diabetes	4·5	4	18	14	1·1	12	10	8
Epilepsy	2	8	24	4	1·6	5	14	2
Hypertension	18·9	1	0	35	11·2	0	0	21
Hypothyroidism	2·2	13	20	3	1·4	6	14	1
Pregnancy	2·5	2	4	30	2·1	0	2	19
Rheumatoid arthritis	2·2	10	21	5	1·6	9	11	1

In the left hand half of the table is the mean and distribution of numbers of patients presenting to 36 full time trainees for the first time for that condition in the first six month period: for some trainees this was only the first part of their attachment. In the right hand half of the table is the same analysis for the subsequent part of the attachment of 21 trainees who went on to complete a total attachment of 11 or 12 months.

Table III shows the percentage of patients with these eight conditions seen once only and twice only by the trainees for the condition in question. The left hand half shows the data from 11 full time trainees on six month attachments, and the right hand half the data from 36 full time trainees on 11 or 12 month attachments.

TABLE III—*Percentage of patients seen once or twice only by the trainee for the condition in question*

Condition	Total No of patients	Percentage of patients seen					
		Six month attachment (n = 11)		Total No of patients	11 or 12 month attachment (n = 36)		
		Once only	Twice only		Once only	Twice only
Asthma	130	69	13	619	50	22
Depression	132	52	15	649	38	21
Diabetes	52	71	9	259	52·5	22
Epilepsy	25	80	16	103	72	18
Hypertension	182	44	25	868	48·5	14·5
Hypothyroidism	10	80	10	91	63	10
Pregnancy	201	62	19	1281	43·5	19
Rheumatoid arthritis	22	68	9	106	45	27

Comparison with trainers

Ten pairs of trainers and trainees counted their total number of consultations in one month for managing the diseases in question. The trainees all had recording efficiency (accuracy) ratings of more than 70%; table IV shows the mean results. The trainers were seeing roughly 50% more patients overall a week than the trainees.

Possible relation to features of training practices

It was important to see if there were features of the training practice that might influence the total number of patients seen and the return rate. Five of the conditions (angina, asthma, diabetes, epilepsy, and rheumatoid arthritis) were grouped together for each individual trainee. The total numbers of patients were plotted on scatter diagrams against