

Do selection scores for general practice registrars correlate with end of training assessments?

IAN DAVISON, SARAH BURKE, JULIE BEDWARD and STEPHEN KELLY

WHAT IS ALREADY KNOWN IN THIS AREA

- Several studies in the UK have evaluated the selection of general practice registrars with respect to competencies and equity.
- These studies have not investigated the link between these selection systems and final outcome measures.

WHAT THIS WORK ADDS

- This study provides evidence that in the West Midlands the selection scores correlate significantly with assessments taken at the end of general practice training.

SUGGESTIONS FOR FUTURE RESEARCH

- Research of this kind should be extended to other cohorts and other areas.
- There will soon be a need to undertake similar studies to assess the success of changes to both the selection and assessment of GP registrars.

Keywords: assessment, general practice registrars, vocational training

SUMMARY

With the move towards centralised selection for general practice training in England, there is great interest in the predictive validity of selection systems. However, to our knowledge, no published studies link selection scores to assessments at the end of training. This study examines how well success in assessments at the end of general practice training is predicted by success in the selection process. Data were collected from the 70 doctors who started general practice registrar training in February 2003 in the

West Midlands, UK. The correlations between selection scores (shortlist, panel interviews and role play) and outcome scores (multiple choice exam, audit submission and video submission) were analysed.

It was found that the multiple choice exam scores correlate significantly with the shortlist ($\tau_b=0.19$, $P=0.036$) and panel interview ($\tau_b=0.23$, $P=0.008$) selection scores, but not with role play. The audit submission scores do not correlate significantly with any of the selection scores. Panel interview ($\tau_b=0.39$, $P<0.001$) and role play ($\tau_b=0.44$, $P<0.001$) scores both

correlate strongly with the video submission scores.

This analysis suggests that candidates who score highly in the West Midlands general practice selection process tend to go on to do well in the multiple choice exam and video assessments at the end of their general practice training. All three elements of this selection process are worth retaining. Further research is required to compare the merits of this selection system with other systems within the UK and internationally.

INTRODUCTION

Since April 2000, the selection processes for general practice (GP) registrars across the UK have undergone critical review.¹ GP directors have responsibility for the organisation and management of a recruitment and training system that is 'fair, open and effective in choosing and developing the doctors who will best provide quality GP services for NHS patients'.² Several regions within the UK have developed competency-based approaches to this selection process. Favourable evaluations of these new systems have been published in terms of the competencies used, changing to a regional system, and equity.³⁻⁷ Whilst these evaluations were favourable, a better measure of the success of a GP selection process would be in terms of the candidates' performance in assessments at the end of GP training. This paper reports on such an investigation.

METHOD

This study analyses selection and outcome assessment data for the 70 doctors who started their GP registrar training in February 2003 in the West Midlands Deanery, UK. The main aim of this study was to investigate how well the outcome assessments were predicted by the selection procedure. That is to say, do candidates who

do well during selection go on to perform better when their GP training is assessed?

Data collection

Two sources of data were used in this analysis: the doctors' scores during the selection process and the results of their outcome assessments.

Details of this selection process have been published elsewhere, but briefly it consists of:⁸

- a shortlist score derived from the candidate's written application
- a panel score based on the candidate's oral responses to two interview panels
- a role play score based on the candidate's performance in a clinical scenario with a trained role player.

The total selection score was calculated by combining these three components.

Outcome assessment included:

- a multiple choice question exam (MCQ) aimed at testing core and emerging knowledge and deeper understanding and application⁹
- a written submission based on an audit project (audit assessment or 'the audit')¹⁰
- a video submission to assess consultation skills (video assessment or 'the video').¹¹

Statistical analysis

In order to correlate the audit with the selection scores we needed to code the audit submissions. A candidate whose first submission was passed by the first examiners was assigned a code of four. If these examiners referred on the submission, and it was passed by the second set of examiners, the code was three. Where the candidate was required to resubmit their audit, a code of two was given if it was passed by the first set of examiners, a code of one if it was passed by the next set, and zero if failed.¹⁰

The video submission is also pass/fail in nature, and GP trainees can choose whether to do the required summative video assessment or the alternative that is part of the membership exam of the Royal College of General Practitioners (MRCGP). The coding used reflects the view that the MRCGP is considered to be the 'gold standard', whilst the summative video assessment assesses a minimal level of competence. So a code of four was used when the MRCGP was passed first time, and a code of three was used if the candidate passed the summative video assessment first time. A candidate who had to resubmit once was coded as two, a code of one was used for a pass after two submissions and two failed submissions were coded as zero.

Results for these audit and video outcomes are non-normal in distribution and contain a very high number of tied ranks. Also, the shortlist scores have a lot of tied ranks. Therefore, Kendall's τ_b has been used to calculate the correlation coefficients.¹² SPSS software was used to analyse the data.

RESULTS

Some of the outcome data for nine doctors were missing because they were taking

longer than normal to complete their training. This was mainly because they had innovative training posts, although one was due to maternity and another due to ill health.

Sixty-four out of the 70 GP registrars (91%) had passed their MCQ exam in the deanery and so were included in the analysis. One doctor had passed elsewhere and five had not yet submitted. Figure 1 displays the distribution of marks on the MCQ. The pass mark for the MCQ varies, so these marks have been slightly adjusted to make them all relative to a pass mark of 70%.

The correlation of MCQ scores with selection scores is shown in Table 1. The shortlist scores weakly correlate ($\tau_b=0.19$, $P=0.04$) with the MCQ. The panel scores correlate a little more strongly ($\tau_b=0.23$, $P=0.008$) with the MCQ exam suggesting that a high score by the panel of assessors tends to predict doing well in the MCQ. The total selection score also correlates with the MCQ ($\tau_b=0.22$, $P=0.01$), but the role play:MCQ correlation is not significant. In summary, success in the MCQ exam is significantly correlated with the total selection score, the panel score and the shortlist score, but not with the role play score.

Of the 65 (93%) GP registrars who had submitted their audit, 39 (60%) were

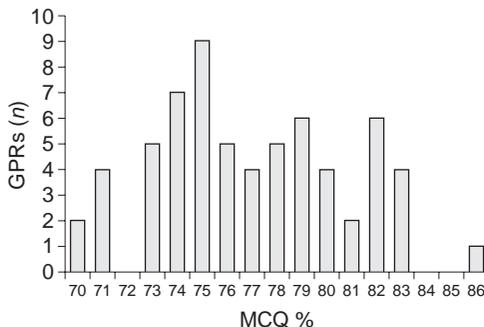


Figure 1 Distribution of MCQ marks

Table 1 Correlations between selection and outcome scores

	Shortlist score/30	Panel score /120	Role play score/60	Total selection score/210	Adjusted MCQ mark	Audit score
Panel score /120	0.146					
Role play score/60	0.077	0.349 (0.000)				
Total selection score/ 210	–	–	–			
Adjusted MCQ mark	0.191 (0.036)	0.229 (0.008)	0.032	0.215 (0.012)		
Audit score	0.012	0.117	0.094	0.106	–0.032	
Video score	0.040	0.389 (0.000)	0.437 (0.000)	0.424 (0.000)	0.155	0.082

2-tail significance levels are given in brackets, where significant.

passed by level 1 markers, and 19 (29%) by level 2 markers. All seven candidates who failed then resubmitted, and five were passed by level 2 markers (audit resubmission), one was passed by national markers (national resubmission), and one failed.¹⁰

The correlations between the GP registrars' success in their audit submissions and the selection scores are also shown in Table 1. The audit scores only correlated weakly and non-significantly with the three parts of the selection process. Therefore, the selection process does not seem to predict which GP registrars will do well at audit.

Sixty-four (91%) of the 70 GP registrars had entered their video for assessment. Forty-three passed the MRCGP video assessment first time, ten who opted to take the summative video assessment passed first time (level 1), seven passed second time (level 2), and four failed both submissions.

Table 1 shows that the video scores correlate strongly with the panel scores ($\tau_b = 0.39$, $P < 0.001$), the role play scores ($\tau_b = 0.44$, $P < 0.001$) and the total selection scores ($\tau_b = 0.42$, $P < 0.001$), although they do not correlate significantly with the shortlist scores.

Table 1 also displays the correlations between the three elements of the selection process. The shortlist scores do not correlate strongly with the other two parts of the selection process; however, there is a high correlation between panel scores and role play scores ($\tau_b = 0.35$, $P < 0.001$), indicating that these two methods of assessment tend to rate the same doctors highly.

A 'good GP registrar' might be expected to do well on all three outcomes. However, Table 1 shows the correlations between these three outcomes as low and non-significant.

DISCUSSION

Interpretation of results

The significant correlations displayed in Table 1 indicate that candidates with high GP registrar selection scores tend to do well in the MCQ and video assessments at the end of training.

The MCQ exam scores correlate with the shortlist and panel selection scores. This may be because these two selection

scores measure attributes such as medical knowledge and understanding required for the MCQ exam. There are two effects that may have reduced the strength of these correlations: first, the MCQ contains a formative assessment element as it can be taken several times; and second, there is a very high pass rate, with only one in 600 GP registrars ultimately failing to complete summative assessment due to the MCQ.¹³ So, the high pass rate and formative nature of the MCQ may mean that 'good' candidates do just enough to pass the MCQ and devote most of their effort to the harder parts of summative assessment. If this is true, these correlations probably underestimate the true value of these selection scores.

The lack of any significant correlation with the audit scores is open to a number of interpretations. First, correlations are unlikely as 60% of candidates are passed by the first examiners, and it is not a discriminating test due to the pass/fail scoring. Second, there are some concerns that audit difficulties may be due to 'problems in teaching and understanding at practice level', so a poor audit may be a reflection on the training practice rather than the trainee.¹³ Even so, it may be that the West Midlands should investigate the testing of skills relevant to audit within its selection procedure.

The correlations of the video scores with the panel and role play selection scores are surprisingly strong given the pass/fail nature of the video assessment and the high pass rates. It seems reasonable to make two conclusions: first, the panel, role play and video assessments are all focusing on similar skills; second, registrars select whether to enter for the harder MRCGP video assessment on the basis of their perception of their competence in these skills. It is particularly interesting to note the high correlation with role play scores which suggests that there is value in retaining this aspect of the selection process.

Recent research suggests that the video is a good test of consultation skills, and

also indicates that failing the video was the most common reason for failing the summative assessment.^{11,13,14} Therefore, the video component could be the hardest element of the summative assessment, and so we speculate that it is the component in which the quality of the candidate is most important. This is a tentative explanation of these relatively high correlations with the video scores.

Strengths and limitations of this study

Construction of effective and efficient selection processes is critically important in all education systems. A recent study has linked the selection system of what was then the Trent Region with formative evaluation after three months' training.¹⁵ However, we believe that no other published study has investigated the link between GP selection scores and assessments at the end of GP training in the UK. Therefore, this study is an important step in developing an evidence base for the selection of GP registrars.

Only one region of the UK was studied in this paper, therefore we are unable to make comparisons with other selection procedures.

The objective of the selection process is to select the best candidates. However, it is clearly not possible to know how well the rejected candidates would have performed in the summative assessment if they had undergone the GP training. Therefore, we are inferring that those rejected candidates would have tended to do less well in the MCQ and video assessments.

Implications

The English National Recruitment Office for General Practice Training was set up in November 2002 (see www.gprecruitment.org.uk), and there are moves towards a national selection system. This paper provides evidence for including panel inter-

views and role play assessments in a future national system.

The summative assessment is a high-stakes examination and training doctors is expensive and time-consuming. It is important, therefore, to select the best possible candidates in the interests of patients and the efficient use of public money, as well as for the doctors themselves. More research of this nature is needed to contribute to the current debate on a national selection procedure for general practice and to aid the development of predictive selection procedures.

References

- Lane P and Sackin P (2003) GP registrar recruitment. *Education for General Practice* **11**: 3–7.
- Department of Health (2000) *The GP Registrar Scheme: vocational training for general medical practice. The UK guide*. Department of Health: London.
- Patterson F, Ferguson E, Lane P, Farrell K, Martlew J and Wells A (2000) A competency model for general practice: implications for selection, training, and development. *British Journal of General Practice* **50(452)**: 188–93.
- Patterson F, Lane P, Ferguson E and Norfolk T (2001) Competency based selection system for general practitioner registrars. *British Medical Journal Career Focus* **323(7311)**: S2–3.
- Burke S, Brown C, Bullock A, Field S and Kelly S (2004) The impact of centralised selection for GP registrars: lessons from the West Midlands. *Education for Primary Care* **15(2)**: 175–82.
- Brown C, Wakefield S and Bullock A (2001) The selection of GP trainees in the West Midlands: audit of assessment centre scores by ethnicity and country of qualification. *Medical Teacher* **23**: 605–9.
- Brown C, Wakefield S and Bullock A (2003) The selection of GP trainees in the West Midlands: second audit of assessment centre scores by ethnicity and country of qualification. *Medical Teacher* **25**: 649–53.
- Wakefield S, Brown C, Bullock A, Field S and Kelly S (2002) Selecting general practitioner registrars: the West Midlands model. *British Medical Journal Career Focus* **325(7354)**: S3–4.
- Royal College of General Practitioners (2002) *The MRCGP Examination*. RCGP: London.
- Evans A, Sloan R, Tinker M and Taylor G (2004) A comparison of results achieved when audit submissions are marked using the COGPED and NPMS methods. *Education for Primary Care* **15(3)**: 344–51.
- Campion P (2004) A new video assessment for the MRCGP exam. *Education for Primary Care* **15(2)**: 153–8.
- Field A (2000) *Discovering Statistics using SPSS for Windows*. Sage: London.
- Murray TS, Attwood M, Lough M, Kelly M and Kelly D (2004) Summative assessment of the competence of GP registrars 1996–2001: reasons for and implications of failure. *Education for Primary Care* **15(2)**: 169–74.
- Swanwick T (2004) The video component of the MRCGP examination: threats to validity. *Education for Primary Care* **15(3)**: 311–27.
- Patterson F, Ferguson E, Norfolk T and Lane P (2005) A new selection system to recruit general practice registrars: preliminary findings from a validation study. *British Medical Journal* **330**: 711–14.

Correspondence to: Ian Davison, Centre for Research in Medical and Dental Education, School of Education, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. Tel: +44(0)121 414 4808; fax: +44(0)121 414 4865; email: i.w.davison@bham.ac.uk

Ian Davison, BA PhD PGCE, Sarah Burke, BA MA, and Julie Bedward, BSc SC PGCE, are Research Fellows at the Centre for Research in Medical and Dental Education, School of Education, University of Birmingham. Stephen Kelly, FRCP FRCGP, is Director of Quality Assurance at the West Midlands Deanery, Edgbaston, Birmingham.

Accepted December 2005