

Do scores in the selection process for vocational general practice training predict scores in vocational examinations?

Steven Lillis MBChB, FRNZCGP, MGP Dip Sports Med¹

¹ Chair, Board of Assessment,
RNZCGP

ABSTRACT

AIM: Selection processes for vocational training are common and are intended to predict future examination performance (predictive validity). Comparison of selection scores with measures of competence during training can provide supporting evidence that both the examination process and selection process are valid and reliable.

METHOD: Selection interview scores for vocational training in general practice were compared to summative examination scores eight months into training between 2003 and 2006.

RESULTS: A moderately strong correlation (Spearman's rho = 0.5) was found between selection interview scores and summative examination scores.

CONCLUSION: The ability of a selection interview process to predict future performance in vocational training is highly dependent on careful design of both the assessment of performance and the selection process. The selection interview process of the RNZCGP compares well with international data.

KEYWORDS: General practice; graduate medical education; professional competence; educational assessment; intake selection

Introduction

Vocational training is expensive and represents a considerable time commitment by the candidate. Training bodies also invest substantial resources in training programmes and have a vested interest in selecting candidates who will succeed in the programme. The selection process for vocational training commonly includes an interview as well as variable collection of information from sources such as referees' reports, academic results etc. The objective of the process is primarily to identify those candidates who are most likely to perform well in the training course or who will perform well later in their professional career (predictive validity).

In New Zealand, vocational training is undertaken by the Royal New Zealand College of General Practitioners (RNZCGP). There is no previously

published research concerning the predictive validity of the RNZCGP interview system. This paper compares candidates' selection scores with scores achieved in a summative assessment (Primex) eight months after commencing the vocational training scheme.

Methodology

Data consisting of an interview score, scores in the written component of Primex and scores in the clinical component of Primex were available for the four years from 2002 to 2006. Incomplete data sets for individual candidates were removed (15), leaving 209 results available for analysis. The underlying structure of both the assessment and the selection process remained unchanged for the four years in which data were available. Therefore the annual data were aggregated.

J PRIMARY HEALTH CARE
2009;1(2):114–118.

CORRESPONDENCE TO: Steven Lillis

Senior Lecturer in
General Practice,
Waikato Clinical School,
Bryant Education Centre,
Private Bag 3200,
Hamilton, New Zealand
slillis@wave.co.nz

The interview process

The interview contains nine sections including a section on referees' reports that are scored by the interviewers. Each interview is allotted one hour and guidelines provided for time limits for each question. Guidelines are also provided on scoring for each question. Three interviewers are used for each candidate; each interviewer scores that candidate separately and, at the end of the interview, a consensus score is achieved by discussion. At least one of the interviewers must be an experienced interviewer who will provide guidance to a novice regarding the interview process.

The assessment

Primex is a two stage assessment; written and clinical. The written component comprises two three-hour examinations containing Multi Choice questions, Extended Matching questions and Key Feature problems. The content of the paper is constructed by active practitioners to ensure that the content has high validity. The clinical component is an eight station Objective Structured Clinical Examination that emphasises communication skills and problem management of direct relevance to the work of a general practitioner. Scores in both written papers and the clinical paper for each candidate were aggregated so that the written component and the clinical component were equally weighted to give a 'combined' score. The internal consistency of the written examinations was reasonable, averaging a Cronbach alpha of 0.8 over the years of data capture. The clinical examinations were consistently scoring around 0.75 over the years for the same time period.

Results

Spearman's rho and p values were calculated for the relationship between interview score and assessment score in the written examination, the clinical examination and the combined examinations.

Two statistical measurements, Pearson's r and Spearman's rho, are commonly used to measure the correlation between two variables. Pearson's r can be used if the data are normally distributed whereas Spearman's rho does not assume normal distribution of data. In this study, the data did not approximate a normal distribution and therefore

WHAT GAP THIS FILLS

What we already know: The validity of both selection processes and of examination processes in vocational training are of considerable interest to training organisations. It would be expected that one would predict the other, yet little research has been undertaken to verify this predicted association.

What this study adds: This research confirms that the selection process as used by the Royal New Zealand College of General Practitioners during the data collection years reasonably predicted future summative examination performance.

Table 1. Interview questions

Interview question	Score
Q1 Firm intention to enter general practice	/10
Q2 Understanding of the nature of general practice	/08
Q3 Readiness for vocational education for general practice	/06
Q4 Commitment to addressing govt priority health areas	/05
Q5 Commitment to professional development and teaching	/03
Q6 Communication skills, ability to work well in a group	/08
Q7 Self preservation and use of support people, own GP	/02
Q8 Flexibility and availability	/04
Referees' reports rating	/04
TOTAL	/50

Table 2. Results of correlations

	Spearman's rho	p
Written examination against interview score	0.446	0.000
Clinical examination against interview score	0.492	0.000
Combined examinations against interview score	0.520	0.000
Written examination against clinical examination	0.310	0.000

p < 0.05 indicates a statistically significant correlation

Spearman's rho was chosen for statistical analysis. A commonly accepted interpretation by Franzblau of the results of both these statistical methods states that less than 0.2 indicates negligible correlation, 0.2–0.4 is a low degree of correlation, 0.4–0.6 a moderate correlation, 0.6–0.8 is marked correlation and more than 0.8 is high correlation.¹

The data reveal a Spearman's rho of 0.52 for the association between the interview score and combined scores of the examination. A slightly weaker association was found for the association between the interview and the individual examina-

Scatterplots of the relationships show:

Figure 1. Interview score against written examination score

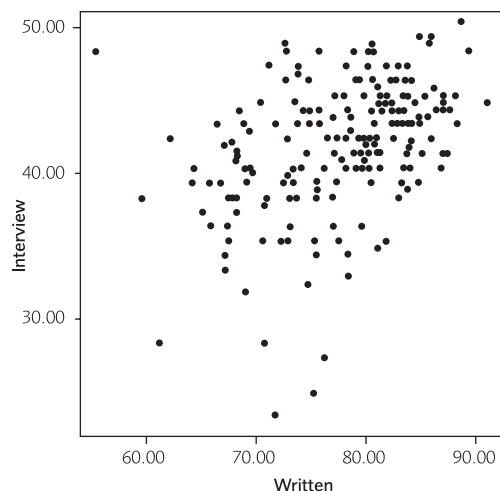


Figure 2. Interview score against clinical examination score

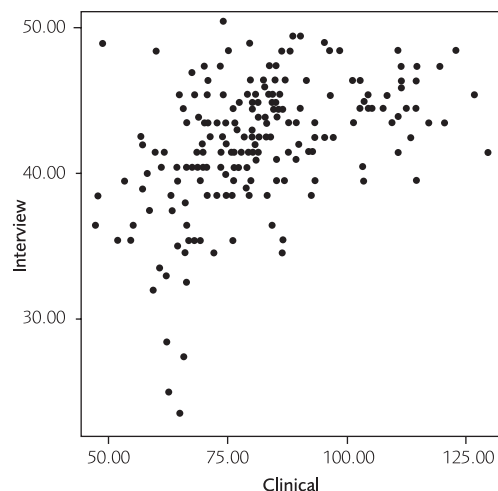


Figure 3. Interview score against combined written and clinical examinations score

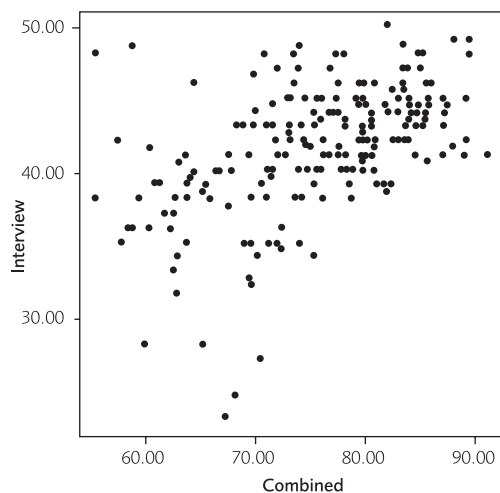
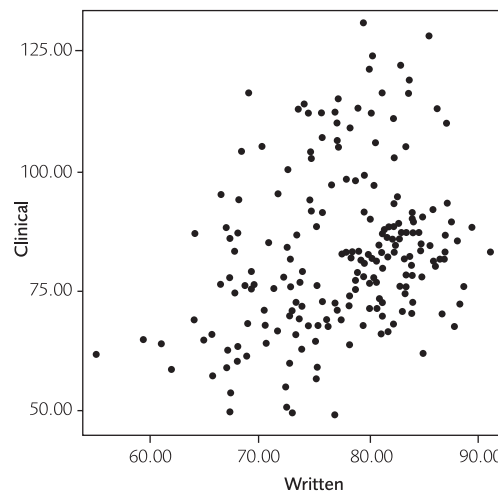


Figure 4. Clinical examination score against written examination score



tion components (0.492 for the clinical and 0.446 for the written examinations). The structure of the interviews did not change in any substantial way during the interval of data collection and the structure of the examination process was also essentially unchanged. It is reasonable, therefore, to use the cumulative data rather than the annual data as a more precise indication of relationship given the limited numbers of candidates.

Hence the relationship between the selection process and a summative assessment taken eight months into the vocational training programme

(14 months post interview) can be considered to be moderate.

Discussion

A review of employment interview research classified interviews according to construct.² At one extreme lies a 'subjectivist-social' approach where the process can be characterised as a social encounter; the roles of interviewee and interviewer are negotiated and the outcome includes better understanding of mutual expectations. The interviewee is an active participant in the process. The

other extreme is an 'objectivist-psychometric' approach where the desired outcome is to achieve a high degree of reliability by controlling for bias and error in interviewer judgement. The interviewee becomes a passive conveyor of information with little or no control over the process. The interview structure reported in this paper strongly reflects an 'objectivist-psychometric' design with high reliability and validity as the objective.

The perplexing issue of measuring performance must also be addressed. Competence demonstrated in knowledge- and skills-based assessments is an indicator of, but does not necessarily predict, a consistent and equivalent level of performance.³ Hence competence can only be considered a surrogate endpoint for measuring performance. As much as it would be ideal to measure performance, the tools required are less than perfect. What is left are the more psychometrically robust measurements of competence and the requirement that we accept the limitations implicit in assessing performance by measuring competence.

A significant body of literature has described the controversies of using selection interviews for admission to medical and health related training courses.⁴⁻⁹ The evidence suggests that interviews are poor overall predictors of both future competence and performance, particularly academic success. There is somewhat less literature concerning the correlation between vocational selection scores and subsequent indicators of performance or competence. The conclusions of predictive validity studies on undergraduate selection processes may not be generalisable to vocational medical training where applicants are older, have considerable medical experience and are the 'survivors' of a long undergraduate course. The purpose of the selection interview may also not be entirely clear; is selection aimed at finding those who will function well as trainees or those who will perform well after training? The two do not necessarily require the same skill sets.

A study of 18 anaesthetics residents found a positive correlation (Pearson's r of 0.48 and p of 0.03) between selection scores and results in an examination taken at the end of the first year of residency training.¹⁰ However, the study failed to find any significant correlation between selection

scores and other measures of performance such as knowledge, judgement, intrapersonal attributes and attitudes, motor skills, pre-anaesthetic assessment and planning and an overall impression. Similarly, a study of 69 paediatric resident medical officers found weak correlation (Pearson's r of 0.28 and p of 0.02) between selection interviews and faculty performance rating, but there was no correlation between the national resident match process and faculty ratings and no correlation between faculty ratings and an in-service examination.¹¹ Selection committee scores for internal medicine residents have been found to moderately correlate with resident final evaluation scores (Pearson's r = 0.52).¹² Clinical performance was found to moderately correlate (Pearson's r = 0.6) with a residency matching programme for 107 obstetrics and gynaecology residents.¹³

The psychological research literature gives guidance on how to construct reliable and valid selection interviews.^{2,14,15} Many of these principles (highly structured interview with identical questions for each candidate, time limits for interview sections, anchored rating scale for each question, multiple interviewers etc.) are features of the selection process reported here. The inclusion of non-academic questions is an important part of the construction process and includes questions designed to measure the 'fit' between the resident and the training scheme. Incorporating multiple factors in the selection process is generally associated with better prediction of competence and/or performance. This could reasonably be expected because the increased number of 'samples' in the interview improves the reliability and the attributes being sampled are reasonably homogeneous.

The 'council of perfection' would suggest that designing a reliable and valid interview process would start with a comprehensive job analysis, the development of specific interview questions and other selection processes based on such an analysis, training of interviewers, careful attention to consistency during interviews and feedback mechanisms by which the performance of the interview process can be monitored. The feedback mechanisms (measures of either competence or performance) must also have acceptable reliability and validity to be meaningful. Both

processes (selection and future assessment) must be constructed with commonality of purpose.

This study also revealed a weak correlation between the clinical and written examinations ($r = 0.31$). The format, content and purpose of the examinations are significantly different in many respects. The knowledge and skills tested are different and are not necessarily generalisable between the two examination processes. Data were selected for the years 2003 to 2006. Changes were made to the interview process beyond 2006 (addition of new interview questions) that would have introduced a confounding variable even accepting that the additional question maintained the objectivist-psychometric position.

The limitation of this study centres around examinations representing a surrogate endpoint for performance as a vocationally registered general practitioner. An examination will, for most candidates, reflect maximal competence. For those few who do not perform well in such circumstances, examination scores may have little bearing on their ability to function as a vocationally trained doctor. Demonstrating maximal competence in an examination does not necessarily relate to consistent high levels of performance. Ideally, both selection scores and examination scores would be correlated with measures of performance as a vocationally registered general practitioner.

Future research should focus on the predictive validity of selection interviews for performance as a general practitioner rather than maximal competence. The predictive validity of Primex against performance would also provide useful information for the development of the examination. In order to achieve these goals, reliable and valid tools for measuring performance will need to be developed or modified from existing tools. Further, there will have to be wide acceptance in the profession of periodic, external and objective measurement of performance.

Conclusion

The ability to predict future performance of a doctor in a specialist training scheme is clearly of significant value, yet historically has been accord-

ed low priority in the educational continuum. This research demonstrates that the predictive validity of selection process for vocational general practice training in New Zealand was moderately high for the years in which data were collected and compares well with vocational training selection processes described in overseas data. To achieve such a correlation, the selection interview must be 'objectivist-psychometric' in design with high reliability and validity. Similarly, the assessment process must also have features of high reliability and validity. Methods of comparing selection processes with vocational post-training performance need to be developed.

References

1. Franzblau A. Primer of statistics for non-statisticians: Harcourt, Brace & World; 1958.
2. Anderson N. Eight decades of employment interview research: A retrospective meta-review and prospective commentary. *Eur J Work Organ Psychol* 1992;2(1):1-32.
3. Rethans JJ, Norcini JJ, Baron-Maldonado M, Blackmore D, Jolly BC, LaDuca T, et al. The relationship between competence and performance: implications for assessing practice performance. *Med Educ* 2002;36(10):901-9.
4. Blanchard J. Board scores and resident performance: Is there a link? *Ann Emerg Med* 2000;36(1):64-7.
5. Borowitz SM, Saulsbury FT, Wilson WG. Information collected during the residency match process does not predict clinical performance. *Arch Pediatr Adolesc Med* 2000;154(3):256-60.
6. Brothers TE, Wetherholt S. Importance of the faculty interview during the resident application process. *J Surg Educ* 2007;64(6):378-85.
7. Goho J, Blackman A. The effectiveness of academic admission interviews: An exploratory meta-analysis. *Med Teach* 2006;28(4):335-40.
8. Kulatunga-Moruzi C, Norman GR. Validity of admissions measures in predicting performance outcomes: the contribution of cognitive and non-cognitive dimensions. *Teach Learn Med* 2002;14(1):34-42.
9. Morris J. The value and role of the interview in the student admissions process: A review. *Medical Teacher* 1999;21(5):473-481.
10. Metro DG, Talarico JF, Patel RM, Wetmore AL. The resident application process and its correlation to future performance as a resident. *Anesth Analg* 2005;100(2):502-5.
11. Peskun C, Detsky A, Shandling M. Effectiveness of medical school admissions criteria in predicting residency ranking four years later. *Med Educ* 2007;41(1):57-64.
12. Fine PL, Hayward RA. Do the criteria of resident selection committees predict residents' performances? *Acad Med* 1995;70(9):834-8.
13. Olawaiye A, Yeh J, Withiam-Leitch M. Resident selection process and prediction of clinical performance in an obstetrics and gynecology program. *Teach Learn Med* 2006;18(4):310-5.
14. Robertson I, Smith M. Personnel selection. *J Occup Organ Psychol* 2001;74:441-472.
15. Campion M, Palmer D, Campion J. A review of structure in the selection interview. *Pers Psychol* 1997;50(3):655-702.

COMPETING INTERESTS

None declared.