

# Predictive validity of a selection centre testing non-technical skills for recruitment to training in anaesthesia

T. C. E. Gale<sup>1\*</sup>, M. J. Roberts<sup>1,2</sup>, P. J. Sice<sup>1</sup>, J. A. Langton<sup>1,3</sup>, F. C. Patterson<sup>4</sup>, A. S. Carr<sup>1</sup>, I. R. Anderson<sup>1</sup>, W. H. Lam<sup>1</sup> and P. R. F. Davies<sup>1</sup>

<sup>1</sup> Anaesthesia Recruitment Validation Group, Department of Anaesthesia, Plymouth Hospitals NHS Trust, Derriford Road, Plymouth PL6 8DH, UK

<sup>2</sup> Institute of Clinical Education, Peninsula College of Medicine and Dentistry, Tamar Science Park, Plymouth PL6 8BU, UK

<sup>3</sup> South West Peninsula Deanery, Tamar Science Park, Plymouth PL6 8BY, UK

<sup>4</sup> Department of Psychology, City University, London EC1V 0HB, UK

\* Corresponding author. E-mail: thomas.gale@phnt.swest.nhs.uk

## Key points

- There is a need for better methods for recruitment and selection of trainees.
- A programme using four stations—structured interview, portfolio review, presentation, and simulation—was evaluated.
- There was good inter-rater agreement.
- Assessment score and subsequent clinical performance in the first year correlated.
- Further work involving the non-appointed candidates is required.

**Background.** Assessment centres are an accepted method of recruitment in industry and are gaining popularity within medicine. We describe the development and validation of a selection centre for recruitment to speciality training in anaesthesia based on an assessment centre model incorporating the rating of candidate's non-technical skills.

**Methods.** Expert consensus identified non-technical skills suitable for assessment at the point of selection. Four stations—structured interview, portfolio review, presentation, and simulation—were developed, the latter two being realistic scenarios of work-related tasks. Evaluation of the selection centre focused on applicant and assessor feedback ratings, inter-rater agreement, and internal consistency reliability coefficients. Predictive validity was sought via correlations of selection centre scores with subsequent workplace-based ratings of appointed trainees.

**Results.** Two hundred and twenty-four candidates were assessed over two consecutive annual recruitment rounds; 68 were appointed and followed up during training. Candidates and assessors demonstrated strong approval of the selection centre with more than 70% of ratings 'good' or 'excellent'. Mean inter-rater agreement coefficients ranged from 0.62 to 0.77 and internal consistency reliability of the selection centre score was high (Cronbach's  $\alpha=0.88-0.91$ ). The overall selection centre score was a good predictor of workplace performance during the first year of appointment.

**Conclusions.** An assessment centre model based on the rating of non-technical skills can produce a reliable and valid selection tool for recruitment to speciality training in anaesthesia. Early results on predictive validity are encouraging and justify further development and evaluation.

**Keywords:** anaesthesia; assessment, educational; education, training; patient simulation; personnel selection

Accepted for publication: 10 June 2010

Recruitment into postgraduate medical training is a widely debated topic, yet there is a lack of strong evidence with which to inform policy.<sup>1,2</sup> A report by the British Medical Association (BMA) in 2006 into selection methods for speciality training<sup>2</sup> stated that there was insufficient evidence to support the unequivocal recommendation of any one particular method. Traditional panel interviews are frequently used in medicine but are subject to error, as judgements are often made 'by committee' rather than truly independent scoring.<sup>3</sup> We describe the design and validation of a new selection centre for recruitment to speciality training in anaesthesia, based on the assessment centre model adopted in other occupations.

Internationally recognized definitions of what constitutes an assessment centre have been developed and refined by practitioners,<sup>4</sup> and research conducted over more than 50 yr supports their acceptability, low adverse impact, and predictive validity across most occupational groups.<sup>5-7</sup> In contrast, the use of assessment centres for medical recruitment is relatively new, although research evidence supporting their use is now emerging.<sup>8-11</sup> Preliminary validation results have been reported for recruitment into general practice,<sup>9</sup> but there have been no predictive validity studies to evaluate assessment centres for postgraduate recruitment to anaesthesia.

In selection to competitive training programmes, it is desirable to distinguish candidates who are excellent or have aptitude for the speciality from those who are merely clinically competent. Non-technical skills are essential for doctors and can distinguish those who demonstrate excellence.<sup>12 13</sup> They are vital for effective functioning in challenging environments and assessment tools have been developed for use in anaesthesia.<sup>14 15</sup> We therefore based our selection centre scoring system on the assessment of candidates' non-technical skills. The aim of the study was to evaluate the reliability, face validity, and predictive validity of this selection centre.

## Methods

Ethical approval for the study was obtained from NHS Cornwall and Plymouth Research Ethics Committee. Candidates attending interview for anaesthesia training posts in the South West Peninsula Deanery during 2007 and 2008 were consented to participate in the study.

An expert panel from the South West Peninsula region identified and agreed the selection criteria. The panel consisted of six Royal College of Anaesthetists' tutors from hospitals in the region, an education director, a senior occupational psychologist, a speciality training director, an academic anaesthetist, a specialist in Human Resources, and an expert in simulation.

The results of a multi-method job analysis study of anaesthetists provided a validated competency model for use in designing the selection criteria.<sup>16</sup> At a 1 day workshop, the expert panel reviewed the competency model and by consensus, identified six criteria to be the most important at the point of selection: (1) achievements, (2) communication, (3) working under pressure, (4) organization and planning, (5) situational awareness/decision-making, and (6) teamworking. The panel judged these selection criteria to go beyond broad clinical competence and to also focus on the large body of literature showing that non-technical skills are essential for practitioners to function effectively in the speciality.<sup>12 14 17</sup>

To design the selection centre, the expert panel was extended to include four additional anaesthetists including

two with skills in simulation and the assessment of non-technical skills plus a manager responsible for post-graduate recruitment across many specialities. During a series of 12 half-day workshops, stations were designed to facilitate assessment of the six selection criteria.

The selection centre comprised a total of four stations, with two stations assessing past behaviour/achievements (*structured interview* and *portfolio review*) and two stations testing the candidates' behaviour during simulated work sample tasks (*presentation* and *simulation*). The centre was designed such that each selection criterion could be assessed in at least two stations using standardized anchored rating scales developed by the expert panel (an example scoresheet for *Teamworking* is shown in Supplementary Appendix A). For each station, two consultant anaesthetists independently rated specific attributes and also awarded a 'global rating score' for overall professional judgement of candidate performance at that station. The *global rating score* also provided a means to veto interviewees who perform particularly poorly or exhibit unsafe characteristics in any particular station (see example *global rating scoresheet* in Supplementary Appendix B). All scores were marked on a scale from 1 (poor) to 4 (excellent). If an assessor awarded 1 on the global rating score at a station, both assessors would confer with the Deanery representative to decide whether the applicant's performance was so poor that a veto should be awarded due to significant patient safety issues such as attitude, probity, or lack of basic clinical knowledge. The overall candidate score was an unweighted sum of the individual ratings of the attributes assessed (Table 1) and was used to make final selection decisions.

All assessors were consultant anaesthetists who received 2 days training in interview and assessor skills for the selection centre. Two pilot runs were performed before the selection centre going 'live', utilizing four trainees who were already in post to evaluate the stations and usability of the scoring system. Each station lasted for 10 min (Table 2).

Reactions to the selection centre were sought using a 12-item feedback questionnaire asking candidates to rate each station on a five-point scale (1, poor; 3, satisfactory; 5, excellent) for (a) *relevance to selection*, (b) *fairness*, and (c)

**Table 1** Blueprint for assessment of candidate attributes in the Selection Centre

Aspects rated	Selection centre stations			
	Interview	Portfolio	Presentation	Simulation
Personal skills/attributes				
Communication	✓	✓	✓	—
Organization	—	✓	✓	—
Working under pressure	—	—	✓	✓
Situational awareness and decision-making	✓	—	—	✓
Teamworking	✓	—	—	✓
Content of portfolio	—	✓	—	—
Global rating	✓	✓	✓	✓

**Table 2** Content of the four selection centre stations

**Structured interview:** Six question areas, designed to reveal behavioural challenges that candidates had experienced in their clinical practice. The interviewers probed candidates with questions exploring their approach to the situation, their interaction with other staff/colleagues, and stimulating reflection on their own performance

**Portfolio review:** The content of each candidate's personal development portfolio was reviewed in conjunction with structured questions eliciting reflection on achievements, career to date, and future professional developmental needs

**Presentation:** Candidates were given 10 min to prepare a short presentation independently on a pre-determined subject. They were advised that the presentation should last 7 min, allowing 3 min for questions by the assessors. Topics were chosen so that this exercise was not a test of knowledge, but a test of how an applicant could deliver a presentation in challenging circumstances with an associated time-pressure for preparation

**Simulation:** Standardized scenarios were produced using medium fidelity simulation (Laerdal SimMan Classic) which incorporated two clinical skills nurses working 'in role' to provide realism in a measured way. A realistic simulated environment was established in existing training rooms without the need for a high cost, high fidelity simulation centre. The scenarios began with a 3 min period for the candidate to assess the manikin's physiological derangement, get into role, apply monitoring, and examine the clinical paperwork prepared specifically for each scenario. The programme then followed a period of deterioration with set criteria to reverse it

*opportunity to demonstrate ability.* The questionnaire was filled out by the candidates immediately after completing their cycle to prevent employment decisions influencing their comments and was completed anonymously to encourage honest responses. The same questionnaire was also used to gather feedback from the assessors with an additional four items exploring the extent to which they judged each station to be (d) *appropriate for selection purposes.*

As part of the validation process, the candidates subsequently appointed consented to participate in a longitudinal study investigating their subsequent performance. Workplace performance was assessed using: (i) in-theatre assessment (ITA) of non-technical skills displayed by trainees while anaesthetizing a particular key index case within 6 months of appointment. Scoring was performed independently by both the consultant anaesthetist and an operating theatre practitioner using the standardized anchored rating scales developed for the selection centre, and (ii) scores derived from trainees' Annual Review of Competence Progression (ARCP) comprising multiple assessments of professional and clinical skills by consultant anaesthetists. Eleven aspects of trainee performance during the first year of appointment were rated on a four-point scale (0, unsatisfactory; 1, weak; 2, typical; 3, very good) by each consultant anaesthetist. Ratings for each aspect were averaged across raters before being combined into a single score for each trainee.

Reliability of the selection centre process was assessed by examining inter-rater agreement between the paired

assessors, internal consistency reliability of the candidates' total scores, and inter-station correlations. Inter-rater agreement was measured using Scott's Pi rather than Cohen's  $\kappa$  since (a) candidates were assessed by different pairs of assessors at individual stations, and (b) assessor scores were recorded anonymously. Owing to the ordinal nature of the rating scale, the Pi statistic was quadratically weighted to allow credit for partial agreement between pairs of raters.<sup>18</sup> Internal consistency reliability of candidates' total scores was measured using Cronbach's  $\alpha$  and relationships between the four stations and the total score were quantified using the Pearson correlation coefficients. Owing to non-normality, station feedback ratings were compared using the Kruskal-Wallis tests with *post hoc* comparisons controlled for multiple testing by the Bonferroni correction.<sup>19</sup>

Predictive validity was assessed using the Pearson correlations between selection centre scores and the two separate measures of workplace performance, the ITA and the ARCP, described above. These correlations are subject to attenuation by (i) imperfect reliability (in both predictor and criterion variables) and (ii) selection (restriction of range). We followed accepted practice by correcting for criterion unreliability and range restriction but not for predictor unreliability.<sup>20</sup>

## Results

In 2007, a total of 143 candidates attended the selection centre and 37 (26%) were appointed to training posts. In the following year, 31 (38%) of 81 candidates were appointed. All 68 appointees consented to participation in the follow-up study.

Inter-rater reliability across the 15 station/attribute assessment points in the first year of the selection centre showed a mean value of Scott's Pi (raw agreement rate) of 0.67 (65%) with a range of 0.55–0.81 (59–71%). A slightly higher mean of 0.70 (70%) was obtained in the following year, with a range of 0.53–0.87 (58–81%). In both years, 13 of the 15 agreement coefficients were more than 0.60, indicating substantial agreement or better.<sup>21</sup>

The internal consistency reliability coefficient (Cronbach's  $\alpha$ ) for the total selection score was 0.88 in 2007, increasing to 0.91 in the second year. Positive correlations between the four station scores and the selection centre total (Table 3) are broadly consistent in magnitude, confirming that the stations form a coherent evaluation of candidate performance. The weaker positive correlations between stations show that although there is some shared variance in station scores, there is no excessive overlap or duplication of assessment.

The overall response rate for the candidate feedback survey was 79% (178/224). In both years, candidates gave approval to the selection tools. Over 70% of ratings were positive (*good or excellent*) in both years and the proportion of negative ratings (*borderline or poor*) decreased from 7.5% in the first year to just 3.8% in 2008. The assessors returned similarly positive ratings. Summary rating statistics for the

**Table 3** Correlation matrix for station scores and selection centre total, 2007/2008. \*\*\*Differs significantly from zero,  $P < 0.001$ ; \*\*differs significantly from zero,  $P < 0.01$ ; \*differs significantly from zero,  $P < 0.05$ 

	Structured interview	Portfolio station	Presentation station	Simulation station
2007 (n=143)				
Portfolio station	0.35***			
Presentation station	0.21*	0.23**		
Simulation station	0.29***	0.16	0.18*	
Selection centre total	0.69	0.64	0.64	0.64
2008 (n=81)				
Portfolio station	0.49***			
Presentation station	0.39***	0.22*		
Simulation station	0.41***	0.37**	0.33**	
Selection centre total	0.77	0.66	0.67	0.78

**Table 4** Mean (SD) candidate and assessor feedback ratings on a five-point rating scale: 1, poor; 3, satisfactory; 5, excellent. \*\*\*Significantly higher scores for relevance, opportunity, and appropriateness of the simulation station (Kruskal–Wallis test,  $P < 0.001$ , plus *post hoc* tests with 0.05 family-wise error rate)

Aspects rated	2007 candidates (n=135)	2008 candidates (n=43)	2008 assessors (n=24)
Relevance to selection			
Structured interview	3.89 (0.86)	4.09 (0.75)	3.78 (0.85)
Portfolio station	3.85 (0.98)	4.17 (0.70)	4.36 (0.68)
Presentation station	3.60 (0.91)	3.93 (0.77)	3.87 (0.95)
Simulation station	4.39 (0.74)***	4.21 (0.91)	4.67 (0.58)***
Fairness			
Structured interview	3.96 (0.87)	4.21 (0.78)	3.91 (0.73)
Portfolio station	3.87 (0.91)	4.29 (0.72)	4.32 (0.61)
Presentation station	3.98 (0.74)	4.01 (0.75)	4.17 (0.72)
Simulation station	4.18 (0.88)	4.05 (0.79)	4.43 (0.60)
Opportunity to demonstrate ability			
Structured interview	3.64 (0.94)	4.07 (0.67)	3.96 (0.64)
Portfolio station	3.69 (1.03)	4.15 (0.79)	3.93 (0.66)
Presentation station	3.71 (0.92)	4.02 (0.80)	4.04 (0.64)
Simulation station	4.11 (0.87)***	4.14 (0.80)	4.71 (0.56)***
Appropriateness to selection			
Structured interview	N/A	N/A	3.78 (0.85)
Portfolio station	N/A	N/A	3.96 (0.69)
Presentation station	N/A	N/A	3.91 (0.67)
Simulation station	N/A	N/A	4.67 (0.58)***

different stations show that the simulation station stood out with regard to 'relevance to selection', 'opportunity to demonstrate ability', and 'appropriateness to selection' (Table 4).

All 68 appointees were followed up in their first year of training. The ITA was completed by 27 (73%) of the 2007 cohort and 18 (58%) of the 2008 cohort, and ARCP scores were obtained for all trainees except one from the 2008 cohort who resigned for personal reasons. In relation to the ITA, there was no evidence of response bias: total scores in the selection centre did not differ significantly between respondents and non-respondents and neither did the gender balance.

The reliability of the ITA score (Cronbach's  $\alpha = 0.87$  for both cohorts) was similar to that of the ARCP ( $\alpha = 0.89$  in 2007 and 0.90 in 2008). There was a reasonable correlation between the selection centre total and both of the subsequent workplace assessment scores (Table 5).

## Discussion

The selection centre performed well in terms of inter-rater agreement, internal reliability, and consistency. All four stations correlated well with overall candidate performance but without evidence of duplication of assessment.

**Table 5** Correlation coefficients between ITA, ARCP, and selection centre scores

Cohort year	Assessment	Sample size	Correlation with selection centre total
2007	ITA	27	0.42
	ARCP	37	0.48
2008	ITA	18	0.34
	ARCP	30	0.33

Consequently, each station had a unique part in the overall process. Assessor and candidate ratings confirmed high face validity of the process. Positive candidate reactions are an important consideration as candidates with positive perceptions of a selection process report stronger intentions to accept a job offer.<sup>5</sup> Early results from longitudinal follow-up demonstrated good correlations with workplace performance 12 months into training.

Inter-rater agreement coefficients for the 15 individual assessment points in our study mostly achieved the 0.60 level indicating substantial agreement or better.<sup>21</sup> The measurement of situational awareness in the interview station ( $r=0.55$  in 2007) and working under pressure in the portfolio station ( $r=0.53$  in 2008) was lower than this threshold. Familiarization with scoring tools, benchmarking of assessors, and scrutiny of individual questions used at each station should help increase inter-rater agreement in future years. Internal consistency reliability of the total selection centre score ( $\alpha=0.88$  in 2007,  $\alpha=0.91$  in 2008) demonstrates a level which is acceptable for high-stakes assessment.<sup>22</sup>

A range of selection methods have been used in recruitment to speciality training including panel interview, psychological profiling,<sup>23</sup> tests of moral reasoning,<sup>24</sup> and the multiple mini-interview (MMI).<sup>25</sup> The MMI has shown good reliability and validity, although most studies have focused on undergraduates.<sup>26–29</sup> Unlike an assessment centre, the MMI does not necessarily include real or simulated job performance tasks or the assessment of attributes in multiple stations. Assessment centres offer candidates *multiple* situations (interview, work simulation, written exercise, etc.) to demonstrate key skills and to be observed by trained assessors.<sup>4</sup> Thus, a fairer (multiple opportunities to perform) and more reliable (multiple observations of key behaviours by multiple observers) assessment can be made. Assessment centres go beyond interviews in that stations can be designed to allow opportunities for candidates to *demonstrate* behaviour relevant to the selection criteria, rather than simply describe what they would do (or think they ought to do) in a given situation.

The incorporation of a global rating score has strengthened the overall reliability of our selection process and has been validated previously in a selection centre for postgraduate primary care training,<sup>9</sup> MMIs for medical school admission,<sup>25</sup> and simulation-based assessment of non-technical

skills.<sup>30</sup> Preliminary results for predictive validity of our selection centre are encouraging, as the correlations between the selection centre overall rating and subsequent work place assessment ( $r=0.33–0.48$ ) compare favourably with those for an assessment centre recruiting to primary care ( $r=0.35$ )<sup>9</sup> and for assessment centres in other occupations ( $r=0.28–0.37$ ).<sup>6, 7</sup>

The BMA report on evidence-based selection<sup>2</sup> concluded that ‘Combining selection methods appears to be the most favourable approach...’. We have incorporated measures of past behaviour (structured interview) and past achievements (portfolio station) with work-related tasks into an assessment centre model for medical recruitment, and shown that non-technical skills are potentially a reliable measure of candidate ability. Non-technical skills underpin the quality of personal performance in challenging situations and their importance has been highlighted in many acute medical specialities<sup>12 14 31 32</sup> as well as high reliability organizations outside healthcare.<sup>33 34</sup> Simulation training in both the aviation and healthcare industries focuses heavily on the importance of non-technical skills which can be assessed during team performance.<sup>17</sup> Simulation is increasingly being used as a tool for assessment of anaesthetists, with evidence of good reproducibility and construct validity<sup>35 36</sup> but has limited use previously in high-stakes situations. Our selection centre is designed to assess particular non-technical skills that are important in anaesthesia. The simulation station, for example, allows us to measure ‘situational awareness’, an attribute that is essential in anaesthesia and may be difficult to assess by other means. Communication skills and the way that anaesthetists interact with patients and other members of the anaesthesia team are equally important.<sup>37</sup>

Our study has highlighted certain non-technical skills that can be reliably assessed in a selection context, and we think that these skills are appropriate. The five skills chosen are part of the UK national person specification for training posts in anaesthesia but have also been identified by studies of anaesthetists’ perceptions of qualities important to clinical practice in their speciality.<sup>16 38</sup> A previous study<sup>38</sup> described professionalism in anaesthesia in categories of humanistic qualities, personal development qualities, and meta-competences. Many of these meta-competences are incorporated as non-technical skills in our study and personal development qualities are assessed in the portfolio station.

This study represents a longitudinal evaluation of the use of non-technical skills for selection and is the first of its kind for postgraduate recruitment to training posts in anaesthesia. We used two outcome measures of workplace-based assessment: one speciality-specific and one generic. Incorporating data from the ARCP report ensured 100% follow-up of appointees who are obliged to take part in this process in order to progress to the next year of training. Nevertheless, our sample sizes are limited since we were unable to follow-up non-appointed candidates in a comparable way due to the regionally based nature of selection and training in the UK. This feature is a common limitation of selection

research<sup>39</sup> but something which could be addressed in the future with larger follow-up studies using nationally standardized recruitment methods.

We have used speciality-specific workplace-based assessments to test the predictive validity of the process, but further work is needed to help develop workplace-based assessment tools which are able to discriminate between trainees' performance reliably. We can then focus on testing the incremental validity of each method used for selection, so that differential weighting can be applied to particular stations or attributes if appropriate. We are planning further longitudinal follow-up and are currently piloting two further stations testing work-related tasks in order to assess the incremental validity of increasing the number of stations used.

In conclusion, a selection centre based on the assessment of non-technical skills shows promise in reliably selecting doctors with an aptitude for anaesthesia. Combining measures of past behaviour/achievements with performance of work-related tasks forms a coherent evaluation of an applicant's ability without evidence of duplication of assessment. Further development and evaluation on a larger scale are required to help refine this recruitment methodology.

## Supplementary material

Supplementary material is available at *British Journal of Anaesthesia* online.

## Acknowledgements

We thank Neil Squires (Recruitment Manager), Martin Beaman (Professor and Postgraduate Dean), and other officers of the South West Peninsula Deanery for assistance in running the selection centres. We thank Cathy Ralph, Justin Phillips, Chris Seavell, and Jonathan Ambler from the South West School of Anaesthesia for help in designing the selection centre structure, and David Adams, Gemma Crossingham, and John Saddler (Head of School of Anaesthesia) for assistance in co-ordinating longitudinal follow-up of appointees. The authors also appreciate the trainees, anaesthetists, and other healthcare professionals who took part in the selection process and subsequent assessment of appointees.

## Conflict of interest

F.C.P. conducts work for the Work Psychology Group, which advises the Department of Health (DH) in the UK on selection and recruitment issues. A.S.C. is seconded to the DH's Moderating Medical Careers team.

## Funding

The work was undertaken by the Anaesthesia Recruitment Validation Group in the South West Peninsula Deanery. Funding was granted as a pilot for selection to acute specialities from the Department of Health (England). M.J.R.'s research post was funded through this grant, as were sessional time allocations for T.C.E.G., P.J.S., and I.R.A. All

opinions in the manuscript are those of the authors only and recommendations or policy from the DH should not be inferred from this work.

## References

- 1 Jefferis T. Selection for specialist training: what can we learn from other countries. *Br Med J* 2007; **334**: 1302–4
- 2 BMA. *Selection for Specialty Training*. British Medical Association Board of Medical Education, 2006
- 3 Cook M. Traditional ways of selecting medical staff. Career focus. *Br Med J* 1998; **316**: classified suppl: 2–3
- 4 International Task Force on Assessment Center Guidelines. Guidelines and ethical considerations for assessment center operations. *Public Pers Manage* 2000; **29**: 315–31
- 5 Thornton GC III, Gibbons AM. Validity of assessment centers for personnel selection. *Hum Resour Manage Rev* 2009; **19**: 169–87
- 6 Gaugler BB, Rosenthal DB, Thornton GC III, Bentson C. Meta-analysis of assessment center validity. *J Appl Psychol* 1987; **72**: 493–511
- 7 Hermelin E, Lievens F, Robertson IT. The validity of assessment centres for the prediction of supervisory performance ratings: a meta-analysis. *Int J Select Assess* 2007; **15**: 405–11
- 8 Mitchison H. Assessment centres for core medical training: how do the assessors feel this compares with the traditional interview? *Clin Med* 2009; **9**: 147–50
- 9 Patterson F, Ferguson E, Norfolk T, Lane P. A new selection system to recruit general practice registrars: preliminary findings from a validation study. *Br Med J* 2005; **330**: 711–4
- 10 Randall R, Davies H, Patterson F, Farrell K. Selecting doctors for postgraduate training in paediatrics using a competency based assessment centre. *Arch Dis Child* 2006; **91**: 444–8
- 11 Ziv A, Rubin O, Moshinsky A, et al. MOR: a simulation-based assessment centre for evaluating the personal and interpersonal qualities of medical school candidates. *Med Educ* 2008; **42**: 991–8
- 12 Glavin RJ. Excellence in anesthesiology: the role of nontechnical skills. *Anesthesiology* 2009; **110**: 201–3
- 13 Smith AF, Greaves JD. Beyond competence: defining and promoting excellence in anaesthesia. *Anaesthesia* 2010; **65**: 184–91
- 14 Fletcher GCL, Flin RH, McGeorge P, Glavin RJ, Maran NJ, Patey R. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. *Br J Anaesth* 2003; **90**: 580–8
- 15 Flin R, Patey R, Glavin R, Maran N. Anaesthetists' non-technical skills. *Br J Anaesth* 2010; **105**: 38–44
- 16 Patterson F, Ferguson E, Thomas S. Using job analysis to identify core and specific competencies: implications for selection and recruitment. *Med Educ* 2008; **42**: 1195–204
- 17 Fletcher GCL, McGeorge P, Flin RH, Glavin RJ, Maran NJ. The role of non-technical skills in anaesthesia: a review of current literature. *Br J Anaesth* 2002; **88**: 418–29
- 18 Fleiss JL. *Statistical Methods for Rates and Proportions*. New York: John Wiley & Sons, 1981
- 19 Siegel S, Castellan NJ Jr. *Nonparametric Statistics for the Behavioural Sciences*. New York: McGraw-Hill, 1988
- 20 Lee R, Miller KJ, Graham WK. Corrections for restriction of range and attenuation in criterion-related validation studies. *J Appl Psychol* 1982; **67**: 637–9
- 21 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74
- 22 Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ* 2004; **38**: 1006–12

- 23 McDonald JSMD, Lingam RPM, Gupta BM, Jacoby JMDP, Gough HGP, Bradley PMA. Psychologic testing as an aid to selection of residents in anesthesiology. *Anesth Analg* 1994; **78**: 542–7
- 24 Self DJ, Baldwin DC Jr. Should moral reasoning serve as a criterion for student and resident selection? *Clin Orthopaed Relat Res* 2000; **378**: 115–23
- 25 Humphrey S, Dowson S, Wall D, Diwakar V, Goodyear HM. Multiple mini-interviews: opinions of candidates and interviewers. *Med Educ* 2008; **42**: 207–13
- 26 Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Med Educ* 2004; **38**: 314–26
- 27 Harris S, Owen C. Discerning quality: using the multiple mini-interview in student selection for the Australian National University Medical School. *Med Educ* 2007; **41**: 234–41
- 28 Lemay J-F, Lockyer JM, Collin VT, Keith A, Brownell W. Assessment of non-cognitive traits through the admissions multiple mini-interview. *Med Educ* 2007; **41**: 573–9
- 29 Reiter HI, Eva KW, Rosenfeld J, Norman GR. Multiple mini-interviews predict clerkship and licensing examination performance. *Med Educ* 2007; **41**: 378–84
- 30 Weller JM, Bloch M, Young S, et al. Evaluation of high fidelity patient simulator in assessment of performance of anaesthetists. *Br J Anaesth* 2003; **90**: 43–7
- 31 Flin RH, Maran NJ. Identifying and training non-technical skills for teams in acute medicine. *Qual Saf Health Care* 2004; **13**: i80–4
- 32 Bion JF, Abrusci T, Hibbert P. Human factors in the management of the critically ill patient. *Br J Anaesth* 2010; **105**: 26–33
- 33 Flin RH, Martin L, Goeters K, et al. Development of the NOTECHS system for assessing pilots' CRM skills. *Hum Fact Aerospace Saf* 2003; **3**: 95–117
- 34 Toff NL. Human factors in anaesthesia: lessons from aviation. *Br J Anaesth* 2010; **105**: 21–5
- 35 Gaba DM, Howard SK, Flanagan B, Smith BE, Fish KJ, Botney R. Assessment of clinical performance during simulated crises using both technical and behavioral ratings. *Anesthesiology* 1998; **89**: 8–18
- 36 Murray DJ, Boulet JR, Kras JF, McAllister JD, Cox TE. A simulation-based acute skills performance assessment for anesthesia training. *Anesth Analg* 2005; **101**: 1127–34
- 37 Smith AF, Mishra K. Interaction between anaesthetists, their patients, and the anaesthesia team. *Br J Anaesth* 2010; **105**: 60–8
- 38 Kearney RA. Defining professionalism in anaesthesiology. *Med Educ* 2005; **39**: 769–76
- 39 Cook M. *Personnel Selection: Adding Value Through People*. Chichester: Wiley, 2004