Discerning quality: using the multiple mini-interview in student selection for the Australian National University Medical School

Susanna Harris & Cathy Owen

OBJECTIVE To describe the development and pilot testing of a set of admissions instruments based on the McMaster University multiple mini-interview (MMI) and designed to assess desirable, non-cognitive characteristics in order to inform final decisions on candidate selection for entry to medical school.

METHODS Community and faculty consultation on desirable, non-cognitive characteristics of medical students informed the development of a 10-station interview. Two stations occurred as part of a group problem-based learning scenario and 8 occurred as individual observations. All interviewers were trained. Interviews were offered to 115 candidates on an academic merit list. Interview performance was used to exclude candidates considered unsuitable, but not to re-order the academic merit list. Admissions decisions were examined in terms of individual interview station performance.

RESULTS This method proved to be an efficient process by which to interview candidates and to determine suitability. Retained and rejected candidates had significantly different total scores and mean scores for each station. Ten independent observations contributed to each decision, without significant interviewer or logistic burden. Candidates reported high levels of satisfaction with the interview process.

CONCLUSIONS Admissions interviews can be streamlined and efficient, yet remain informative. A longitudinal study is in progress to evaluate the value of the admissions processes in predicting successful graduation to medical practice.

KEYWORDS *school admission criteria; *schools, medical; *education, medical, undergraduate; interviews/*standards; communication; motivation; problem solving; attitude; Australia.

Medical Education 2007; ***41****: 234–241*
doi:10.1111/j.1365-2929.2007.02682.x

Medical Education Unit, Australian National University, Canberra, Australian Capital Territory, Australia

Correspondence: Professor Cathy Owen, Medical Education Unit, Australian National University, Building 42, Canberra, Australian Capital Territory 0200, Australia. Tel: 00 61 2 6125 3973; Fax: 00 61 2 6125 8877; E-mail: cathy.owen@anu.edu.au

INTRODUCTION

Medical schools generally agree that although high academic achievement is the minimum requirement for success in medical school, it is not sufficient to admit students solely on the basis of academic performance. Measures of superior academic performance, such as high grade point averages (GPAs) and high standardised entrance examination scores, have been shown to be good predictors of pre-clinical performance, but function 'much more poorly on prediction of clinical performance in medical school and probably not at all with the ultimate goal of competence as a physician'.[1] The importance of assessing a candidate's non-cognitive characteristics in selection for admission to medical school is reflected in the common statements about such qualities in accreditation guidelines such as those of the General Medical Council in the UK, the American Board of Medical Specialties, and the Accreditation Council for Graduate Medical Education in the USA, and CanMEDS 2000[2] in Canada.

Although the personal interview is the most commonly used instrument for assessing a candidate's non-cognitive traits, it is considered to have limited reliability. Several factors contribute to the low

Overview

What is already known on this subject

Admissions decisions are complex and are improved by the use of the multiple mini-interview (MMI).

What this study adds

A method for determining a set of non-cognitive characteristics on which to base individual MMI stations is described. Pilot data are presented.

Suggestions for further research

Longitudinal examination of admissions screening based on non-cognitive characteristics at admission and relationships with junior doctor outcomes is required to determine the usefulness of this approach.

reliability of standard one-to-one or panel interviews, even assuming that the interview is highly structured. The high correlation of ratings among a panel interviewing candidates together may result from working and talking together. More importantly, the personal interview is context-specific. This means that interview scores may be determined less by a candidate's characteristics than by the context in which the interview is held, including the make-up of the panel. Although panels can have good inter-rater reliability, the repeatability of the panel's judgements, even with the same candidate, may be poor. We previously used a structured interview with a 4-member panel (and a group first aid activity) to screen candidates. Analysis, however, showed a lack of discrimination power between candidates (especially for the labour involved) driving the development of this approach.

Eva *et al.* designed a new admissions tool, based on the objective structured clinical examination (OSCE), called the multiple mini-interview (MMI).[3] At the heart of this new tool is the notion that reliability rises with an increase in sampling, thus allowing users of the MMI to gain an accurate picture of each person's strengths and limitations. In a series of articles, Eva *et al.* described the development and piloting of the MMI, which they demonstrated to be feasible, acceptable and reliable.[4] The instrument was tested using both graduate volunteers and real candidates for admission. McMaster University, in Hamilton, Ontario, is continuing to study the performance of the MMI and its predictive value.[5]

As in many medical schools, small-group learning – especially that focused on problem-based learning (PBL) – is central to our programme. The ability to participate in group activities was therefore seen as a legitimate requirement in prospective students. Others' experience with group activities as an admissions screening tool has shown that using clearly defined rating criteria, which consider both qualitative and quantitative observations, improves the reliability and validity of a group screening task.[6] We therefore elected to add observations from a group encounter as extra stations to our MMI.

The Australian National University (ANU) Medical School participates in a national admissions scheme, using academic performance data from each candidate's previous degree and a common entry examination. This information is used to rank applicants, after which the upper 120 applicants are offered interviews for 80 study places. Previous work has highlighted the risk of placing too much weight on a limited interview sample from each candidate.[1,3] On this basis, this school decided to restrict the use of the admission interview to the exclusion of unsuitable candidates, rather than the revised ranking of suitable candidates. This defines the admission interview as a tool suitable for identifying outlier candidates (those most 'unsuitable'), but not capable of making fine distinctions between groups of 'suitable' candidates.

To determine suitable candidates, we needed to establish the preferred non-cognitive characteristics valued by our community in entry-level medical students. Q methodology was chosen as the means of identifying the traits most valued.[7] This technique was chosen as the assessment and statistical method for the study because it was developed to incorporate qualitative and quantitative research traditions. Q methodology can reveal the subjectivity in a situation and, although it was initially used in personality assessment, it has been applied in a range of psychological investigations.[8] The method allows a quantitative evaluation of the opinion of individuals about topics of common concern. This leads to a composite of opinions that may be aggregated into

viewpoints.[9] Each participant prioritises statements on an agree–disagree continuum using a forced choice model. Analysis reveals common opinions about the examined subject.

METHODS

Ethical approval was obtained from the institutional ethics committee. Statements describing possible non-cognitive characteristics were developed from detailed analysis of previous traditional candidate interviews, consideration of accreditation guidelines and focus groups with previous interviewers (including people from the community, health facilities and the university). This process identified 47 summary statements of non-cognitive characteristics that may be useful in the identification of candidates for medical school.

A further convenience sample (including medical students, early graduates, health academics, clinical health workers and administrative staff) was then approached to sort these non-cognitive characteristics. They were asked to assume that candidates already met the academic standards required for admission. Each was given 47 statements on small cards to sort according to the characteristics they felt were more or less important in an incoming medical student. Statements could be rated from −5 (strongly disagree) to +5 (strongly agree). The active area on the Q sort board, onto which the statements were sorted, took the shape of an inverted triangle, forcing participants to prioritise their choices, as shown in Fig. 1.

Each participant's finished ranking of the statements is known as a 'sort.' Each sort was entered into SAS software (Version 9.1) for analysis. Factors were extracted using principal components analysis and rotated using orthogonal varimax rotation. Factors with eigenvalues > 2 were then selected, which produced 6 factors. Items were deemed to load on a factor if their rotated loadings were ≥0.4 as an absolute value. The identified factors were described and informed the development of the admission interview stations. The nature of the interview and pilot data are presented.

Sorts from 105 participants were collected and analysed. Six factors (representing groups of participants with common views) were identified (Table 1).

The factors are described as follows.

1. Love of medicine and learning: participants valued the qualities of curiosity and enthusiasm about medicine and lifelong learning.
2. Groundedness: participants valued people who were not only overly intellectual or ambitious; they did not consider that the capacity to think laterally was important in isolation.
3. Self-confidence: participants valued the capacity for self-confidence combined with some familiarity with health care systems.
4. Balanced approach: participants valued candidates who were able to take a comprehensive approach to a problem while maintaining a sense of proportion.
5. Mature social skills: participants appreciated insightful, reflective candidates with appropriate social skills but did not believe that active listening skills in isolation were sufficient.
6. Realism: participants valued life experience and a realistic outlook as valuable, non-cognitive qualities in prospective medical students but did not value the interpersonal qualities of humour, empathy and being a group contributor.

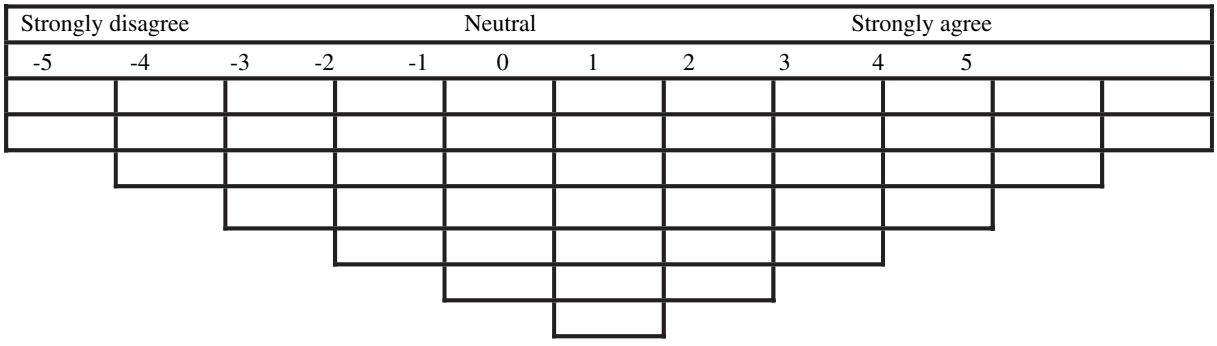

Figure 1 Q-sort board layout.

Table 1 Rotated factor loadings for each identified factor

	Statement	*Factor 1*	*Factor 2*	*Factor 3*	*Factor 4*	*Factor 5*	*Factor 6*
17	Demonstrates enthusiasm for medicine	0.75362					
30	Demonstrates curiosity and enthusiasm in a medical area	0.64610					
38	Shows curiosity and commitment to lifelong learning	0.53474					
31	Is not excessively motivated by personal ambition		0.75749				
34	Is not over analytical or intellectualising		0.57672				
13	Can think laterally about information given		– 0.65106				
37	Is confident			0.78237			
46	Is familiar with health care system			0.58862			
26	Maintains a sense of proportion in the face of a challenge				0.70182		
44	Has a broad comprehensive approach to issues				0.66051		
21	Demonstrates appropriate social interaction					0.76343	
16	Demonstrates insight and the ability to reflect on one's own situation					0.44937	
11	Listens actively					– 0.60460	
41	Has had experience abroad						0.74098
8	Has a realistic view of life as a doctor						0.39546
29	Uses humour and goodwill to work alongside others						– 0.39317
2	Demonstrates an interest in and empathy for people						– 0.48322
14	Contributes to discussion						– 0.53994

These 6 factors informed the consequent development of the 10 independent MMI stations. Two observations were made within the context of a group activity (discussed further below), and 8 observations were carried out in an MMI cycle. The 8 interview stations included:

1. Giving instructions. (Rationale: displaying confidence, technical communication and appropriate social skills, dealing with frustration, maintaining a sense of proportion in the face of the task.)
2. Taking instructions. (Rationale: displaying confidence, technical communication and appropriate social skills, dealing with frustration, maintaining a sense of proportion in the face of the task and having a realistic outlook.)
3. Emotional communication. (Rationale: demonstrating mature social skills and a realistic perspective.)
4. Problem solving. (Rationale: demonstrating the ability to take a comprehensive approach to a problem while maintaining a sense of proportion.)
5. Resilience and maturity. (Rationale: demonstrating life experience and a realistic outlook in dealing with problems.)
6. Enthusiasm for medicine. (Rationale: exploring curiosity and enthusiasm about medicine and lifelong learning.)
7. Ethics. (Rationale: demonstrating a grounded perspective and an awareness of ethics as an issue.)
8. Awareness of common issues in medicine (using rural medicine as a focus to demonstrate some familiarity with health care systems.)

Documentation included information for the interviewer explaining the purpose of the station, prompts for the interviewer, and a rating matrix for station-specific attributes. The record also included a space for note taking, an area to note issues of concern for further discussion, and a 7-point global rating. The station-specific rating matrix was used to support the global rating and had key criteria marked with an asterisk.

For illustration, the station designed to test problem solving is presented. The personalised interview record form included a summary of the purpose of the station and a standard introduction. In this case the interviewer announced that he or she would read out a description of a scenario and that there were no right or wrong answers, and encouraged the candidate to think creatively and broadly. The interviewer then read the scenario. This is an example:

> 'The government wants to track citizens across the county in order to maintain public safety in the face of the growing threat of terrorism. How would you advise the government to do this?'

The interviewer was then given a series of follow-up prompts specific to the station, designed to challenge the candidate and elicit the broadest range of responses. As the interview proceeded, the

interviewer rated the candidate on his or her performance relative to specific station attributes, using scale terms of 'unsatisfactory', 'borderline', 'satisfactory' and 'excellent.' Some attributes were asterisked for mandatory discussion if they were not rated as satisfactory, such as when a candidate was unable to maintain a reasonable demeanour during a communication station. In the space for observations, the interviewer was able to note aspects of the candidate's response. If a candidate gave cause for concern, this was noted as an issue for further discussion. At the end of the interaction, the interviewer assigned the candidate an overall score on a scale of 1–7, where 1 = unsatisfactory, 3 = borderline, 5 = satisfactory and 7 = outstanding.

All interviewers were trained in this interview technique. The role and purpose of the admission interview were clarified and all stations were reviewed. Interviewers were familiarised with all stages of the process by experiencing a mock round of interviews (both as candidates and interviewers) with clarification and debriefing afterwards. Interviewers were drawn from 4 categories: community members; recent medical graduates; university academics, and health facility staff (including rural health staff).

Each station was 5.5 minutes long. Changeover time amounted to 30 seconds. A single interviewer was assigned to each station and remained there for the entire round of 8 candidates. Each interviewer rated each candidate on a personalised interview record.

A bell announced the end of each station and candidates were encouraged to move promptly to their next station. Each candidate started at any position in the cycle. All interviews were audio-recorded after informed consent had been obtained from the candidates.

The last 2 stations occurred in the context of a group session. The group activity was a non-medical PBL session, conducted by an experienced PBL tutor and rated separately by 2 trained silent observers (focused on either qualitative or quantitative observations). The PBL tutor was instructed to guide the group of 8 or 9 candidates through the problem, to encourage them to appoint a scribe, develop an action plan, and allocate tasks to each group member, as if the group would reconvene to solve the problem presented by the trigger scenario. The qualitative observer was asked to record key moments in each candidate's participation that reflected the nature of that person's performance. The quantitative observer was asked to record how frequently each candidate made a positive or negative contribution to the flow and productiveness of the group process. Ratings were made independently. The tutor was encouraged to keep the group focused on the task, as is usual. In addition, the tutor was asked not to intervene in group dynamics too soon, but to allow the candidates' genuine characteristics to emerge clearly enough to be noticed by the observers. All trigger or prompt materials were changed after each candidate round to prevent cross-contamination between interview cycles.

A candidate who was rated unsatisfactory in any of the asterisked station attributes, or who had a notation in the section for discussion for any station or group activity, or who was given an overall rating ≤ 4 on ≥ 2 observations, was considered of questionable suitability and was discussed by the group of interviewers at the end of the cycle. Admission suitability was then determined by consensus. Unsuitable candidates were removed from further consideration. Suitable candidates remained on the merit list. Interview performance did not reorder the merit list.

RESULTS

Trained raters completed all interviews. A total of 115 candidates were interviewed, which required 87 interviewers (each contributing several hours of labour). Demography by outcome is shown in Table 2.

Of those interviewed, 97 candidates were deemed satisfactory and 18 were excluded on the basis of their interview from further consideration for admission. Unsatisfactory candidates were those who did not rate as satisfactory in > 7 of 10 stations, confirmed by discussion with all interviewers. The proportion of women steadily increased throughout the process. The gender balance in the complete sample versus those ultimately enrolled was compared in a 2-by-2 table using chi-square analysis ($P = 0.062$). This difference, although not significant, approaches significance and deserves examination in later cohorts. Any systematic difference is likely to be multifactorial as students may be satisfactory at interview yet elect not to proceed to enrolment.

On the global scale of 1–7 for each station, retained candidates' mean scores ranged between 5.04 and 5.56, with standard deviations of 0.869–1.097. When candidate scores were added together, the retained candidates' total combined scores were higher than those of the excluded candidates. Figure 2 illustrates

Table 2 Candidate demographics

				Age (years)							
	Total	*Female*	*Male*	*< 26*		*26–35*		*> 35*			
Group	*n*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*
All	325	208	64	117	36	214	66	90	28	21	6
Interviewed	115	79	69	36	31	91	79	22	19	2	2
Satisfactory	97	69	70	28	29	76	78	19	20	2	2
Enrolled	81	59	73	22	27	59	73	19	23	3	4

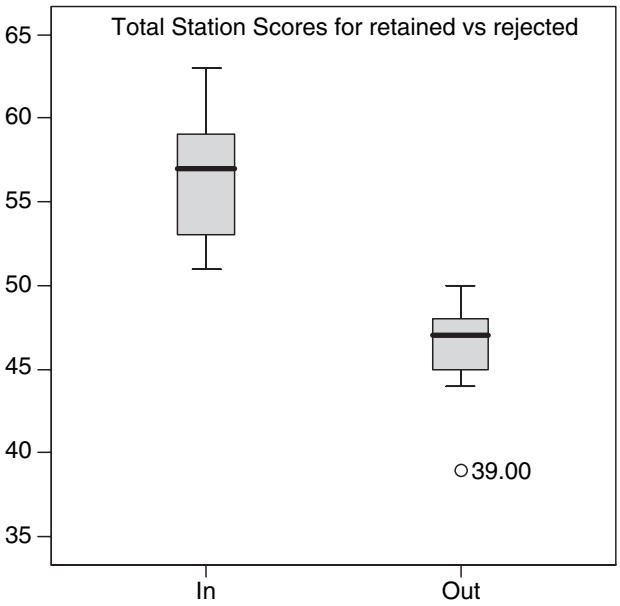

Figure 2 Comparison of total combined scores for retained and rejected candidates.

the clear separation between the combined scores of both groups.

Figure 3 shows the mean (and 95% confidence intervals) for each group for each station. Only 3 stations show any overlap in results, namely, stations 1, 2 and 5. Figure 4 illustrates the small overlap between retained and excluded total scores, which we attribute to some quieter candidates scoring lower on PBL activities. When those candidates were presented for discussion (because they had ≥3 scores of ≤4), if 1 or 2 of those scores came from PBL observers, depending on the discussion, those candidates were often retained. Overall, 17 candidates obtained qualitative PBL ratings of ≤4, 9 of whom were retained; 15 candidates overall achieved quantitative PBL ratings ≤4, 6 of whom were retained. Both PBL observers scored 3 of the retained candidates low. Discussion in favour of retaining a low-scoring candidate focused on the candidate's ability to develop greater confidence and make a significant contribution in group learning with experience.

Of those 18 candidates excluded, at least 5 candidates failed 8 of the 10 stations. Seven of the 18 rejected candidates were marked unsatisfactory by quantitative PBL group observers and 8 by qualitative observers. Nine candidates failed station 7 (ethics); 11 failed station 6 (enthusiasm for medicine and recognition of one's own limitations), and 12 failed station 4 (problem solving). These 3 stations (which stopped the largest numbers of excluded candidates) represent significant non-cognitive characteristics, in particular a lack of emotional commitment to the field, the inability to view one's shortcomings realistically, and mental inflexibility leading to overly concrete thinking.

DISCUSSION

The purpose of this admission interview was to exclude candidates who were believed unsuited to the ANU Medical School programme. The Q sort proved a useful method for determining this school's particular set of desirable non-cognitive characteristics; however, they may not be suitable for every medical school. At this school, interview data were considered a priori to be of insufficient rigour and reliability to warrant using the information to reorder a merit list for placement allocation. Although unsatisfactory candidates were often very unsatisfactory (perhaps in 8 of 10 stations), the small overlap in scores between retained and excluded candidates suggests this decision was sound.

The purpose of using the combination of group activity and the MMI was to gather as many independent datapoints for each candidate as was feasible in order to reduce 'the degree to which candidates are selected based on their chance assignment to a compatible interview team'.[3] One could argue that the MMI presents a fragmented view of a candidate.

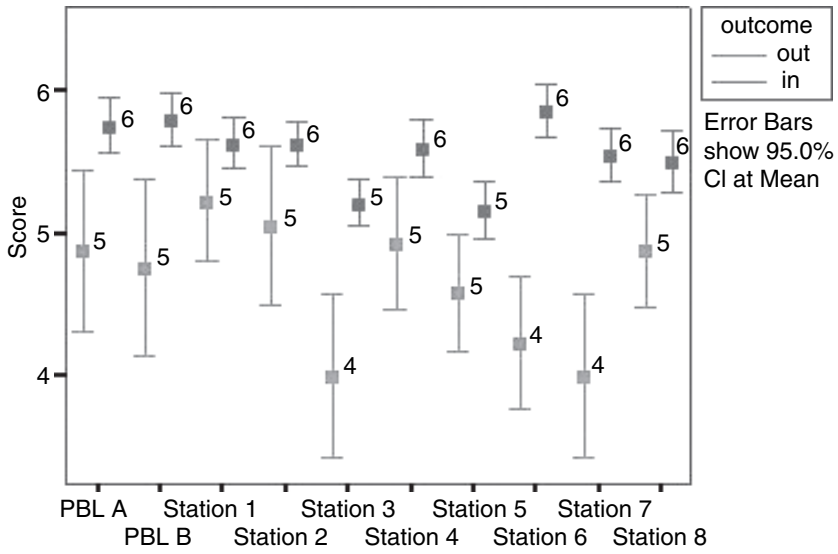

Figure 3 Mean for each station (and 95% confidence intervals) of retained (black) vs. rejected (grey) candidates.

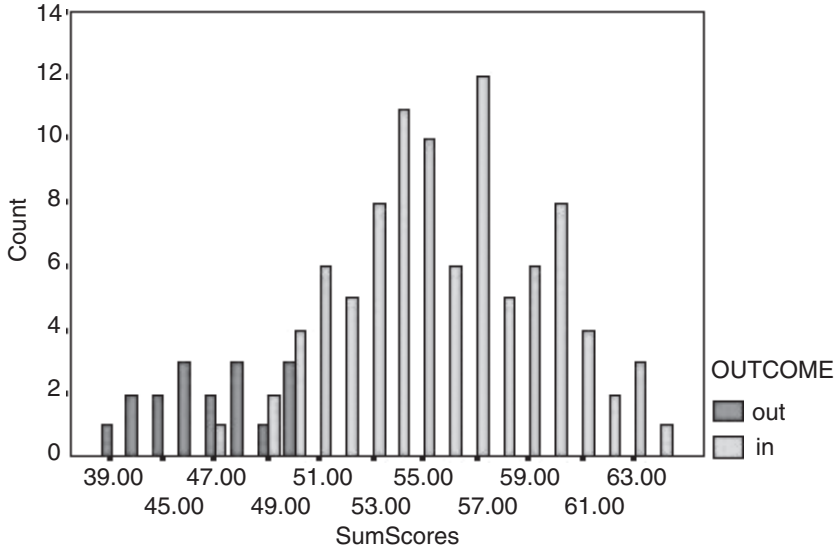

Figure 4 Comparison of accepted vs. rejected candidate total scores.

However, both the PBL activity and the discussion of questionable candidates at the end of each round provided a more holistic counterbalance than the station scores alone. Frequently, quiet candidates obtained poorer scores in PBL than their more voluble colleagues. The interviewing panels retained candidates who were quiet but made pithy contributions and rejected candidates who were marked down for true negative behaviours (such as dominating the group at the expense of others).

Most candidates were rated as satisfactory. More women than men entered the medical school. They were over-represented in the application cohort and their proportion steadily increased. Although this gender imbalance is not yet significant, it is approaching significance and requires following over future years. Women may be more adept at these types of activities, but it would be concerning if there were systematic bias based on gender. Final enrolment figures are determined by positions available (number and type of placements) and candidate preference. More candidates are suitable than there are places available. Without conducting a study of those candidates who met the criteria for exclusion, we will never know whether the excluded candidates represent the 'right' exclusions. In feedback, however, some excluded candidates revealed they attended the interview because of parental pressure and did not wish to study medicine.

All interviewers underwent several hours of training in the entire admissions process, in both the PBL activity and all other stations. This allowed them to understand how the whole admissions protocol worked and on what basis each candidate was judged,

and gave us the advantage of being able to move individuals between functions, allowing for last-minute cancellations. An ancillary advantage of the new admissions procedure is that it required fewer interviewers, was more economical and was in some ways logistically simpler. These interviews took 3 days to carry out and interviewers enjoyed the experience. The process allowed the opportunity to make an independent observation of each candidate and negated any need to manage a panel of perhaps disparate interviewers.

Candidates reported they enjoyed the process. They appreciated that each interviewer saw them afresh and was oblivious to earlier errors. As candidates attended in half-day batches of 24 students, there was also an opportunity to welcome them warmly to the school and deal with any queries as a group. Many candidates commented on the positive impact of this and enjoyed meeting and working with fellow candidates. Informed feedback to excluded candidates has, to date, been well received and used to reapply (rather than to challenge the admissions process).

The new ANU admissions procedures are more effective and more efficient than the previous combination of conventional panel interviews and outdoor group activity. The new procedures also appear to have produced more precise, understandable methods of eliminating unsuitable candidates for admission. These assessment methods have facilitated feedback to unsuccessful candidates, many of whom wish to re-present.

We are committed to tracking the efficacy of these procedures longitudinally. There are opportunities for fruitful cross-institutional and cross-cultural collaborations around the further development of MMI stations.

Contributors: both authors made substantial contributions to the study conception and design, and acquisition, analysis and/or interpretation of data. Both authors contributed to the writing of the article and approved the final version.

Acknowledgements: the authors acknowledge Judy Searle and Jane McHarg, Peninsula Medical School, who generously shared details of their problem-based learning group procedures, Kevin Eva, McMaster University, who kindly discussed ideas and approaches, and Mike Jones, University of Sydney, who gave advice on the statistical approach.

Funding: none.

Conflicts of interest: none.

Ethical approval: not applicable.

REFERENCES

1 Glick SM. Selection for entry to medicine and specialist training. *Med Teach* 2000;**22** (5):443–7.
2 Eva KW, Reiter HI. Reflecting the relative values of community, faculty and students in the admissions tools of medical school. *Teach Learn Med* 2005;**17**:4–8.
3 Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Med Educ* 2004;**38**:314–26.
4 Eva KW, Reiter HI. Where judgement fails: pitfalls in the selection process for medical personnel. *Adv Health Sci Educ Theory Pract* 2004;**9**:161–74.
5 Eva KW, Reiter HI, Rosenfeld J. The ability of the multiple mini-interview to predict pre-clerkship performance in medical school. *Acad Med* 2004;**79** (10 Suppl):40–2.
6 Searle J, McHarg J. Selection for medical school: just pick the right students and the rest is easy. *Med Educ* 2003, 2003;**37**:458–63.
7 Brown SR. Q methodology and qualitative research. *Qual Health Res* 1996;**6** (4):561–7.
8 Stephenson W. *The Study of Behaviour: Q Technique and its Methodology*. Chicago: University of Chicago Press 1953;8–10.
9 Brown SR Q. Methodology and qualitative research. *Qualitative Health Res* 1996;**16**:561–7.

Received 14 September 2005; editorial comments to authors 2 December 2005; accepted for publication 12 October 2006