original paper

An admissions OSCE: the multiple mini-interview

Kevin W Eva, Jack Rosenfeld, Harold I Reiter & Geoffrey R Norman

CONTEXT Although health sciences programmes continue to value non-cognitive variables such as interpersonal skills and professionalism, it is not clear that current admissions tools like the personal interview are capable of assessing ability in these domains. Hypothesising that many of the problems with the personal interview might be explained, at least in part, by it being yet another measurement tool that is plagued by context specificity, we have attempted to develop a multiple sample approach to the personal interview.

METHODS A group of 117 applicants to the undergraduate MD programme at McMaster University participated in a multiple mini-interview (MMI), consisting of 10 short objective structured clinical examination (OSCE)-style stations, in which they were presented with scenarios that required them to discuss a health-related issue (e.g. the use of placebos) with an interviewer, interact with a standardised confederate while an examiner observed the interpersonal skills displayed, or answer traditional interview questions.

RESULTS The reliability of the MMI was observed to be 0.65. Furthermore, the hypothesis that context specificity might reduce the validity of traditional interviews was supported by the finding that the variance component attributable to candidate–station interaction was greater than that attributable to candidate. Both applicants and examiners were positive about the experience and the potential for this protocol.

DISCUSSION The principles used in developing this new admissions instrument, the flexibility inherent in the multiple mini-interview, and its feasibility and cost-effectiveness are discussed.

KEYWORDS education, medical undergraduate/*standards; college admission tests/*standards; cost benefit analysis.

*Medical Education 2004; **38**: 314–326*
doi:10.1046/j.1365-2923.2004.01776.x

INTRODUCTION

Because many medical schools, particularly in North America, have very low rates of attrition, one could argue that the admissions procedure is the most important evaluation exercise conducted by a school. Perhaps as a result of this, there is considerable controversy regarding how best to select individuals from pools of highly qualified applicants. A recent review of the evidence for and against the effectiveness of multiple admissions tools used to select students in the health science professions led to the conclusion that 'preadmission GPA [grade point average] is clearly the best predictor of academic performance.'[1] Still, in addition to academic achievement, health sciences programmes value non-cognitive variables such as interpersonal skills, integrity and professionalism. It is less clear that current assessment tools are capable of predicting ability in these domains. Typically, some form of interview is used; by the early 1980s, 99% of medical programmes in the USA were found to use the interview as part of the admissions process,[2] as were 81% of physiotherapy programmes and 63% of occupational therapy programmes.[3] A more recent survey suggests there has been little change in these proportions; Nayer reported that 99% of US medical schools and 83% of US physiotherapy programmes use interviews.[4]

While the face validity of the interview remains strong, evidence of its effectiveness is more equivocal. Interrater reliability estimates vary widely, from 0.14

McMaster University, Hamilton, Ontario, Canada

Correspondence: Kevin W Eva, Department of Clinical Epidemiology and Biostatistics, Programme for Educational Research and Development, T-13, Room 101, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada. Tel: 00 1 905 525 9140 (ext. 27241); Fax: 00 1 905 572 7099; E-mail: evakw@mcmaster.ca

Key learning points

Health sciences programmes have an ethical responsibility to ensure that reliable and valid admissions tools are utilised.

A multiple sampling approach to assessment has been used by the evaluation community to overcome context specificity. A similar principle could be flexibly adopted for admissions tools.

Increasing the number of interviewers should have the effect of diluting the degree to which candidates are selected based on their chance assignment to a compatible interviewer team.

Global rating scales are capable of providing reliable judgements of candidate performance when averaged across multiple mini-interviews. Candidates and examiners alike viewed the MMI as an acceptable admissions protocol.

The MMI requires fewer examiner hours than traditional interviews that take place in front of a panel of interviewers. In addition, the blueprinting required for MMI station generation can provide a stimulus to prompt admissions committees to collectively make explicit the qualities for which they desire to select.

to 0.95, but this inconsistency might largely be an effect of variability in the way in which interviews are administered;[5] structured formats (i.e. standardised questions with, sample answers provided to interviewers) tend to yield higher rates of reliability and validity than do unstructured formats.[6,7] However, even these reliability estimates may be artificially inflated by:

1 the interview team having access to academic information on candidates,[8,9] and

2 non-verbal communication (which is, admittedly, often unintentional) between members of the interviewing team.

As a result, despite acceptable interrater reliability in some cases, a candidate's score may still be attributable, in large part, to chance. A lucky candidate who is randomly assigned to a like-minded, 'easy' interviewer who influences the rest of the interview panel will score highly, whereas an identical, but less fortunate candidate who is randomly assigned to an incompatible, 'hard' interviewer who influences the rest of the interview panel will score poorly.[10] Other biases that have been shown to impinge upon the personal interview include both the interviewers' backgrounds[6,8,11] and the interviewers' expectations.[6,12] In fact, Harasym *et al.* found that interviewer variability accounts for 56% of the total variance in interview ratings.[12] Such strong biases are unacceptable (and unethical) for an assessment tool that is intended to examine the characteristics of the candidate, not the interviewers.

However, it is not simply interviewer bias that limits the generalisability of interview scores. Many of the problems with the personal interview might be explained, at least in part, by the possibility that the personal interview is yet another domain that is plagued by context specificity.[13] Decades of research have indicated that many of our cognitive 'skills' are highly dependent on context.[14,15] In other words, our performance is commonly less determined by 'trait' (the stable characteristics of the individual) than our intuitions suggest, and more determined by the 'state' (the context within which the performance is elicited). For example, an individual's ability to problem solve or communicate effectively when discussing the impact of the magnetic compass on the modern world will not predict with great certainty that individual's ability to problem solve or communicate effectively when discussing the detrimental effect of monopolies on the world's economy.[16] Consistent with this possibility, Turnbull *et al.* showed that, although interrater reliability within the oral interview certification examinations used by the Royal College of Physicians and Surgeons of Canada was high, the generalisability across interview sessions was low, thereby lowering the overall test reliability.[17] As a result, a single interview may not provide an accurate, generalisable portrayal of a candidate's true abilities even though interrater reliability may be improved by standardising the questions asked and training the interviewers. Multiple topics might be raised within an interview, but this may still represent a small sample of possible responses by the candidate and an interviewer's impressions of each response may not be independent of one another.

Similar realisations led many clinical programmes to adopt the objective standardised clinical examination (OSCE) and other 'multiple sample' approaches to assessing clinical competence.[18] The critical insight was that it is necessary to broadly sample an individual's competencies in order to gain an accurate

picture of that individual's strengths and weaknesses. The current paper will first outline the development of an innovative admissions protocol – the multiple mini-interview (MMI) – that is intended to take advantage of this lesson in the context of student admissions and, second, report results from 2 studies of this protocol performed at McMaster University. In testing this innovation, it was necessary to make many decisions based solely on educated intuition. As a result, we make no claims at this point regarding the optimal use of the MMI, but instead present our logic and reasoning with the hope that some of our assumptions and expectations will be further tested in the future.

THE MULTIPLE MINI-INTERVIEW

First and foremost, it should be noted that the term 'OSCE' has been used in the title of this article simply to orient the reader to the protocol that has been developed for the MMI. Like the OSCE, the MMI is intended to consist of a large number of short stations, each with a different examiner. The MMI is not, however, objective. Nor is it clinical. Research on both the clinical reasoning exercise[19,20] and the OSCE[21,22] has shown that subjective ratings can be reliable and valid estimates of an individual's abilities. As a result, we do not view the subjective nature of the interview process itself to be a limiting feature of this admissions tool. Furthermore, we have carefully avoided developing stations that require clinical knowledge in an effort to prevent biasing the process in favour of health sciences students/personnel.

In contrast to what it is not, the MMI is an OSCE-style exercise consisting of multiple, focused encounters. It is intended to assess many of the cognitive and non-cognitive skills that are currently assessed (inadequately) by the personal interview. Its specific advantage is that multiple interviews should dilute the effect of chance and interviewer/situational biases. Unlike traditional interviews, we can ensure that the ratings assigned to the multiple points of discussion are given independently because interviewers engage the applicants in separate rooms.

While the term 'interview' has been maintained, one of the intended benefits of this protocol is the flexibility with which stations can be developed. For any given station, the examiner might be an interviewer or an observer. As an example, a station on ethical decision making, such as station 1 (see Appendix) can consist of a discussion between candidate and interviewer. Obviously some part of the rating assigned by the interviewer will be influenced by the candidate's ability to communicate effectively, but stations that are intended to tap into communication skills more directly can also be developed. For example, communication skills stations might consist of 'interviews' conducted with a 'simulated patient' while the examiner acts as an observer. Station 3 (see Appendix) is one such station in which the candidate is told s/he has to pick up a colleague to fly to a conference only to discover upon entering the room that the 'colleague' has developed a fear of flying as a result of the September 11th tragedy. The observer rates the candidate based on the communication skills and empathy observed during the interaction between the candidate and 'colleague'. This flexibility in station development reduces the likelihood that candidates will benefit from preparing and rehearsing responses to specific questions. Instead of asking the usual historical questions (e.g. Why do you want to become a doctor?), candidates must respond spontaneously to the presented situation. Undoubtedly, candidates will still prepare and rehearse responses, but it will be more difficult to predict the types of questions one will be asked if a database of stations is developed to sufficient size.

If a programme does desire to query applicants regarding their life history, traditional interview stations can be used in which the interviewer allows the candidate to discuss whatever personal experiences, challenges or beliefs s/he would like the admissions committee to recognise. Similarly, if a programme desires to use the interview, in part, as a recruitment exercise, then a station can be assigned for this purpose without fear of impinging upon the rest of the interview process.

For the remaining stations, specific interview topics can potentially be drawn from any subject ranging from art history to zoology. In fact, an anticipated secondary advantage of this new protocol lies in its potential to draw interviewers from diverse academic and community areas and allow them to assess topics that are consistent with their domain of expertise. We opted to focus our test stations on 4 domains that are not considered to be comprehensive, but are considered to be vital for a career in the health sciences:

1 critical thinking;

2 ethical decision making;

3 communication skills, and

4 knowledge of the health care system.

To assess the suitability of potential stations, we decided that candidates should not be expected to possess specialised knowledge. For example, they should not be expected to know details of a medical condition. Rather, stations should be developed in such a way that they allow candidates to display an ability to think logically through a topic and communicate their ideas effectively. In addition, as a simple heuristic, we viewed any question that had a definitively correct answer to be inadequate. That is not to say that some answers are not better than others, but rather that the interviewers should not be searching for a specific catch phrase or a specific opinion.

Ten stations were developed. The 'Instructions to Applicants' for each station are shown in the Appendix. In addition to these instructions, we developed roughly a page of background information (i.e. a description of the issues and intent of the station) as well as a list of potential points of discussion (i.e. arguments and reasoning in favour of both sides of the issue) as aids for the interviewers.

EXPERIMENT 1: PILOT STUDY WITH GRADUATE STUDENT PARTICIPANTS

A study group of 18 graduate students and 12 interviewers were recruited broadly from the Faculty of Health Sciences to mount a pilot test of the MMI. In addition, an actress was recruited from the standardised patient (SP) programme at McMaster University to play the role of Sara (the 'colleague' in station 3). Stations 1–4, 6 and 7 in the Appendix were used, with 2 interviewers assigned to each station.

As in an OSCE, separate rooms were used for each station. Posted to each door was a card with the 'Instructions to Applicants', as shown in the Appendix. In addition, as this was not intended to be a memory task, the same information was included on a card inside the interview room so that the candidate could refer back to it if s/he desired to do so. Each station lasted 8 minutes and was followed by a 2-minute interval during which interviewers completed standardised evaluation forms and candidates prepared for the subsequent station. The evaluation forms requested interviewers to rate each of the candidates using 7-point scales on:

1 communication skills;

2 strength of the arguments raised;

3 suitability for the health sciences, and

4 overall performance.

Table 1 reports the variance components and G-coefficient that indicates the reliability of the test. The overall test generalisability (i.e. the reliability of the average of all 12 ratings) was found to equal 0.81. Table 1 also reports the results of a D-study performed to determine the optimal combination of stations and raters, assuming that 12 observations can be collected. In general, it appears that increasing the number of stations has a greater impact on the reliability of the test than increasing the number of raters within any given station, thereby supporting the hypothesis that context specificity plagues the traditional interview.

EXPERIMENT 2: UNDERGRADUATE MD PROGRAMME CANDIDATES

Methodology

Participants

All applicants ($n = 396$) who were offered an interview by McMaster University's undergraduate medical programme were sent a letter inviting them to participate in an admissions research study. The letter stressed that their participation (or lack of participation) would in no way influence their chances of being accepted to the medical programme and offered candidates $40 in an attempt to make it clear that this initiative was completely separate from the regular admissions process. A total of 182 candidates responded affirmatively, of which the first 120 candidates whose schedules coincided with participation in one of 12 prearranged research sessions were selected. Three sessions were run sequentially during each of the 4 interview days, with a 40-minute break for examiners between sessions. All candidates were allowed to participate only after completion of the regular admissions protocol. Three candidates backed out due to illness, resulting in a total sample size of 117; 2 of these left before completing a post-MMI survey.

Of the 115 who completed the post-MMI survey, 91 (79%) had a science background, 65 (56%) were female, and the average age was 25.87 years (range 19–47 years).

Interviewers were recruited broadly from the Faculty of Health Sciences, the students currently in the

Table 1 Summary of effects, estimated variance components, the G-coefficient and results of the D-study indicating expected reliability for combinations of 12 observations (Experiment 1)

Effect	d.f.	MS	Estimated variance
Candidate	17	9.946	0.742
Station	5	2.914	0.000
Interviewer within station	6	12.095	0.614
Candidate * station	85	2.070	0.514
Candidate * interviewer within station	102	1.041	1.040

G-coefficient

$$\sigma^2_{(candidate)}/(\sigma^2_{(candidate)} + (\sigma^2_{(candidate * station)}/6) + (\sigma^2_{(candidate * interviewer\ w/in\ station)}/12)) = 0.81$$

D-study

$$\sigma^2_{(candidate)}/(\sigma^2_{(candidate)} + (\sigma^2_{(candidate * station)}/n_{(station)}) + (\sigma^2_{(candidate * interviewer\ w/in\ station)}/n_{(interviewer\ w/in\ station)})$$

	$n_{(station)}$	$n_{(interviewer\ w/in\ station)}$	G
1 station, 12 interviewers	1	12	0.55
4 stations, 3 interviewers within each station	4	3	0.77
6 stations, 2 interviewers within each station	6	2	0.81
12 stations, 1 interviewer within each station	12	1	0.85

medical programme, and the community at large (including McMaster University's Human Resources Department). From the surplus of individuals who volunteered to participate, we selected 40 (10 per day) based on their willingness to volunteer for an entire day. Evaluators were mostly drawn from the Faculty of Health Sciences, but 8 students and 2 members of the Human Resources Department also participated. The list of health sciences volunteers included representation from rehabilitation sciences, nursing, biochemistry and medicine.

In addition, 6 actors were recruited from the SP programme at McMaster University to play the roles of Sara (the 'colleague' in station 3) and Tim (the 'BMW driver' in station 8). Each actor participated in one of the sessions on each of the 4 days.

Procedure

The procedure was identical to that of the pilot study with the following exceptions:

1. all 10 stations reported in the Appendix were used;
2. only 1 interviewer was assigned per station, and
3. as a result of the high correlations among the 4 evaluation questions used during the pilot study, we opted to ask evaluators to simply 'score the applicant's overall performance on this station'.

Results

Scores

The average scores the 117 candidates received across 10 MMI stations ranged from 3.2 to 6.55, with a mean of 5.02 (standard deviation = 1.46).

The effects of gender and session were examined using ANOVA and revealed no differences between males and females ($F_{1,106} = 0.139$, $P > 0.7$) and no

drift in the ratings assigned across 'time of day' ($F_{2,106} = 0.048$, $P > 0.9$). The means are shown in Table 2. These 2 factors accounted for so little variance that their inclusion in the ANOVA had little effect on the reliability of the test.

Reliability analyses

To determine the reliability of the test as a whole, a candidate × station ANOVA was performed. Table 3 reports the variance components and illustrates an overall test generalisability (i.e. the reliability of the average of all 10 ratings) equal to 0.65. None of the stations correlated with any other station greater than r = 0.370. Furthermore, the variance attributable to the candidate–station interaction was 5 times greater than that assigned to the candidates themselves, further supporting the hypothesis that context specificity negatively impacts on traditional interviews.

Correlation with other measures

The MMI scores did not correlate highly with any of the other admissions tools currently used by McMaster's admissions protocol. The correlations between the MMI and the existing admissions tools[23] – personal interview, simulated tutorial, undergraduate grade and autobiographical sketch – were r = 0.185, r = 0.317, r = − 0.227 and r = 0.170, respectively. These numbers are consistent with the correlations between other pairs of tools, which averaged r = 0.056. Despite these low correlations and the fact that the MMI data were not available to the admissions committee, those who were admitted to the undergraduate MD programme received significantly higher scores on the MMI (mean = 5.30/7) than those who were not (mean = 4.83/7; F = 6.97, $P < 0.01$).

Post-MMI surveys

Table 4 illustrates the responses given by the candidates regarding their views of the experience. These responses generally indicate that candidates were quite positive about the MMI. In addition, candidates were asked 3 open-ended questions. In response to the question: 'What do you believe to be the greatest benefits of using the MMI?', many commented on the opportunity to recover from poor stations and the

Table 2 Mean scores as a function of gender and time of day (session)

Session	Female	Male	Total
11.00–13.00	4.92	4.81	4.87
13.30–15.30	5.10	5.09	5.10
16.00–18.00	5.08	5.05	5.07
Total	5.04	4.98	5.02

Table 3 Summary of effects, estimated variance components and the G-coefficient (Experiment 2)

Effect	d.f.	MS	Estimated variance
Candidate	111	4.959	0.322
Station	9	13.697	0.107
Candidate * station	999	1.721	1.721

G-coefficient

$\sigma^2_{(candidate)}/(\sigma^2_{(candidate)} + (\sigma^2_{(candidate\ *\ station)}/10) = 0.65.$

belief that the 'MMI should provide a more balanced view of the applicant's skills and experiences'. Positive comments were also recorded regarding the opportunity to maintain a dialogue with the interviewer and the opportunity to 'solve and discuss REAL PROBLEMS [sic]'.

Candidates were also asked the questions: 'Are there any improvements you would like to see made before the MMI is implemented?' and 'What do you believe to be the greatest weaknesses of the MMI?' Their responses to these focused primarily on logistical issues, such as including 'a chair between stations', lengthening the amount of time for each interview (most often suggested as lengthening to 10 minutes) and 'allow[ing] for some discussion at the end, [to provide an] opportunity to go back to a point not adequately covered'. Some commented that the MMI would allow for a 'shorter interview day', but that a 'break half way through would help'. Others noted the lack of an opportunity to reveal group skills – a domain that could potentially be built into future iterations of the MMI.

Table 5 illustrates the responses given by the interviewers regarding their views of the experience. Some examiners commented on the process being fun, but tiring. To combat this, 1 examiner suggested rotating the interviewers throughout the day so that a different station was assessed during each rotation. The benefit of this change would have to be weighted against the cost of lessening the examiner-identified benefit of an increased 'ability to set a standard for expected responses' and the improved 'consistency of comparing responses' that develops from seeing a large number of candidates work through the same station.

Interestingly, in contrast to the comments offered by some candidates, examiners tended to suggest that 8 minutes was more than enough time to get a sense of the candidate's performance. In general, the most consistent comment, raised by approximately a quarter of respondents, was that the examiners would have liked more training beforehand, potentially in the form of including more information and a longer list of potentially relevant questions in the preparatory package received by all examiners.

DISCUSSION

There are 4 issues that need to be considered when evaluating the efficacy of any assessment protocol:

1 reliability;

2 validity;

3 feasibility, and

4 acceptability.

The reliability of the MMI has now been shown to be in an acceptable range (0.65–0.81) across 2 studies, using graduate student volunteers and actual applicants to the undergraduate medical programme. While adequate, this reliability might be further improved with examiner training. The low correlations between various admissions tools including the MMI is consistent with the hypothesis that context specificity impacts upon admissions protocols, thereby further promoting the need for a tool that adopts a multiple observations approach analogous to that provided by OSCEs when assessing clinical competence.

The data collected from this larger study will allow us, in future, to gain insight into whether or not the MMI is a more valid predictor of performance in medical school than the personal interview. Of the 117 individuals who sat the MMI, 48 have entered medical school at McMaster. The performance of these individuals will allow us to compare MMI scores to in-programme performance and eventually to longterm measures such as the licensing examination.

The blueprinting process undertaken for the generation of stations was intended to maximise the content validity of the MMI. We selected 4 domains that are thought to represent important, non-cognitive characteristics for success in the health sciences. We advocate that specific schools and specific programmes within the schools that consider implementing the MMI engage in a similar process, determining the characteristics they value before creating MMI stations. This blueprinting technique might then ensure an optimal match between the curricular tenets of the programme and the characteristics of the individuals accepted into the programme.

That being said, even the most reliable and most valid of admissions exercises will not be useful if they do not prove to be feasible and cost-effective. In fact, the issue of cost-effectiveness ranks high among the primary assaults that have been launched against the use of personal interviews.[5] For McMaster's medical programme, approximately 400 applicants are interviewed annually, each of whom requires an hour of interview time (30 minutes for

Table 4 Candidate responses to post-MMI survey

Question	Adjectives used on scale 1	3	5	7	Mean (SD)
1 Do you believe that you were able to present an accurate portrayal of your ability?	Definitely not	Not really	Somewhat	Definitely	5.64 (0.85)
2 Compared to the actual interview, do you think the MMI would cause candidates more or less anxiety?	A lot more	A little more	A little less	A lot less	3.80 (1.48)
3 Would the use of the MMI stop you from applying to McMaster?	Definitely not	Not really	Somewhat	Definitely	1.36 (0.79)
4 Were the instructions given before the MMI adequate to prepare you for the experience?	Definitely not	Not really	Somewhat	Definitely	5.87 (1.04)
5 Were the instructions given before each station clear enough?	Definitely not	Not really	Somewhat	Definitely	5.84 (0.95)
6 Do you think any of the interviews required specialised knoswledge?	1 None	4 Somewhat		7 A lot	
(a) Station 1 (Placebo)					3.57
(b) Station 2 (Aspartame)					4.23
(c) Station 3 (Air Travel)					1.82
(d) Station 4 (Deterrent Fees)					3.35
(e) Station 5 (Why do you want to be …)					1.67
(f) Station 6 (Circumcision)					4.04
(g) Station 7 (Class Size)					3.10
(h) Station 8 (Parking Garage)					1.54
(i) Station 9 (Preferential Admission)					3.08
(j) Station 10 (Past Experiences)					1.63
Average					2.80
7 How difficult was each interview?	1 Easy	3 Somewhat easy	5 Difficult	7 Very difficult	
(a) Station 1 (Placebo)					3.92
(b) Station 2 (Aspartame)					4.09
(c) Station 3 (Air Travel)					4.00
(d) Station 4 (Deterrent Fees)					3.15
(e) Station 5 (Why do you want to be …)					2.27
(f) Station 6 (Circumcision)					4.36
(g) Station 7 (Class Size)					3.70
(h) Station 8 (Parking Garage)					3.29
(i) Station 9 (Preferential Admission)					3.27
(j) Station 10 (Past Experiences)					2.23
Average					3.43
8 Was the time available for each station appropriate?	1 Too little time	4 Well timed		7 Too much time	
(a) Station 1 (Placebo)					4.00
(b) Station 2 (Aspartame)					3.77

Table 4 Continued

(c) Station 3 (Air Travel)	4.69
(d) Station 4 (Deterrent Fees)	3.89
(e) Station 5 (Why do you want to be …)	3.34
(f) Station 6 (Circumcision)	4.03
(g) Station 7 (Class Size)	3.91
(h) Station 8 (Parking Garage)	5.23
(i) Station 9 (Preferential Admission)	4.29
(j) Station 10 (Past Experiences)	3.41
Average	4.06

Table 5 Interviewer responses to post-MMI survey

Question	Adjectives used on scale 1	3	5	7	Mean (SD)
1 Do you believe that you were able to develop an accurate portrayal of the candidates?	Definitely not	Not really	Somewhat	Definitely	5.68 (0.90)
2 Compared to a more traditional interview, do you think the MMI would be more or less difficult to administer (from the point of view of an examiner)?	A lot more	A little more	A little less	A lot less	4.73 (1.49)
3 Were the materials provided before the MMI adequate to prepare you for the experience?	Definitely not	Not really	Somewhat	Definitely	6.11 (0.88)
4 Were the instructions given to candidates before your station clear enough?	Definitely not	Not really	Somewhat	Definitely	6.20 (0.88)

the interview and 30 minutes for scoring and a break). There are 4 people on each interview team, so each personal interview requires 4 person hours per applicant; the entire interview programme therefore requires 1600 person hours in total. Of these, 550 are typically faculty hours, the cost of which amounts to an estimated $27 500 per annum. The use of other non-cognitive tools, particularly the simulated tutorial, increases total interviewer time to about 1800 hours and faculty cost to about $32 000. A 10-station MMI (with 10 minutes per station) could be run for only 2 person hours per candidate (including a 20-minute break for all examiners). Assuming the same ratios of faculty versus community personnel, this would require 275 faculty hours at a cost of $13 750 per annum. These values could potentially be reduced even further if it is determined that 10 minutes per station is not required or if fewer stations are used (although the disadvantage of this latter strategy will be poorer reliability). The cost will be increased slightly by the use of SPs, with the absolute value of the increase depending on the

number of SP stations used. As one final note on the feasibility of implementing the MMI, most health sciences programmes have considerable experience with mounting OSCEs. This expertise can potentially be used to make the transition from a personal interview to the MMI as smooth as possible.

Finally, a note on the acceptability of admissions tools. As the MMI and personal interview do require more human resources than simple reliance on grades, it is important that the individuals who are asked to act as interviewers are willing to participate in the process. Similarly, all programmes want to attract the best candidates and, as a result, want to avoid deterring potential candidates from applying because of a unique admissions process. The results from the post-MMI surveys, reported in Tables 4 and 5, suggest that acceptance of the MMI should not be a problem. Interviewers in the pilot study were most concerned about the experience being more tiring than the personal interview, because a single person is responsible for each interview. Addressing this concern might require adjusting the protocol, increasing the number or length of breaks, or changing some other aspect of the process. It should be kept in mind, however, that an equal number of interviewers reported the exercise to be 'fun' and entertaining. The longterm balance of these countering perspectives can only be determined in the future. The candidates and pseudo-candidates in the 2 studies also seemed to enjoy the process and reported (albeit after the event) that they would not shy away from applying to McMaster and that they would be caused no more anxiety if the MMI were to be implemented as an admissions tool in place of the traditional personal interview.

CONCLUSION

The ability of the personal interview to select the candidates who are most likely to succeed in the health sciences has been repeatedly called into question.[1,5] In an effort to overcome the problem of context specificity, we have developed a multiple mini-interview protocol that consists of a series of short OSCE-style stations. Preliminary results suggest that the MMI can be a highly reliable tool; predictive validity data are still forthcoming. The anticipated strengths of the MMI are 6-fold:

1 it allows multiple samples of insight into a candidate's abilities;

2 it dilutes the effect of chance and examiner bias;

3 stations can be structured so that all candidates respond to the same questions and interviewers receive background information a priori;

4 admissions directors have a great deal of flexibility in that stations can be designed with a blueprint of the qualities they would like to select for in mind;

5 candidates can feel confident that they will be given a chance to recover from a disastrous station by moving onto a new, independent interviewer, and

6 fewer resources might be required.

This latter point is only potentially important – even if it turns out that the MMI requires the same amount of resources or slightly more, these resources might at least be better spent on a tool that can prove itself more capable of selecting the highest quality candidates. Health sciences programmes do, after all, have an ethical obligation to do everything in their power to make appropriate and accurate admissions decisions because these decisions will have a large impact on the quality of health care received by society. We do not claim at this point that the MMI is the solution to this problem, but the findings of the current pair of studies is cause for optimism.

CONTRIBUTORS

All 4 authors contributed to the development of the model MMI, to the formulation of the research strategy, and to the development of test stations. KWE recruited the candidates, performed the data analyses and wrote the manuscript. JR assessed the feasibility of the MMI, recruited examiners and oversaw many of the logistical issues. HIR had the inspiration to create an admissions OSCE. The paper was proofread and revised by all authors.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the 52 individuals who volunteered a day (or night) of their time and effort by acting as examiners in 1 of the 2 studies; to Jane Bennett, Lori Johnston and Annette Schrapp, of the undergraduate medical education office at McMaster, each of whom applied

their vast OSCE experience toward organising the logistics of the evening and keeping the process on track; to Alkesh Patel, who helped develop the questionnaires used and surveyed the participants after their experience in the MMI, to Sara Cymbalisty, who recruited and trained the simulators and acted as a simulator herself in the pilot study, and to Adam Norman for assisting in the development of the ethical decision making stations. We would also like to thank the 200 candidates and pseudo-candidates who offered to participate.

FUNDING

We are grateful for the financial support provided by the office of the Associate Dean Education, on behalf of the Faculty of Health Sciences, McMaster University.

REFERENCES

1 Salvatori P. Reliability and validity of admissions tools used to select students for the health professions. *Adv Health Sci Educ* 2001;**6**:159–75.

2 Puryear JB, Lewis LA. Description of the interview process in selecting students for admission to US medical schools. *J Med Educ* 1981;**56**:881–5.

3 Dietrich MC, Crowley JA. A national study of student selection practices in the allied health professions. *J Appl Health* 1982;**11**:248–60.

4 Nayer M. Admission criteria for entrance to physiotherapy schools: how to choose among many applicants. *Physiother Can* 1992;**44**:41–6.

5 Morris JG. The value and role of the interview in the student admissions process: a review. *Med Teacher* 1999;**21**:473–81.

6 Edwards JC, Johnson EK, Molidor JB. The interview in the admission process. *Acad Med* 1990;**65**:167–75.

7 Powis DA. How to do it – select medical students. *BMJ* 1998;**317**:1149–50.

8 Elam CL, Andrykowksi MA. Admission interview ratings: relationship to applicant academic and demographic variables and interviewer characteristics. *Acad Med (Suppl)* 1991;**66**:13–5.

9 Shaw DL, Martz DM, Lancaster CJ, Sade RM. Influence of medical school applicants' demographic and cognitive characteristics on interviewers' ratings of non-cognitive traits. *Acad Med* 1995;**6**:532–6.

10 Mann WC. Interviewer scoring differences in student selection interviews. *Am J Occup Ther* 1979;**33**:235–339.

11 Elam CL, Johnson MM. An analysis of admission committee voting patterns. *Acad Med* (Suppl) 1997;**69**:72–5.

12 Harasym PH, Woloschuk W, Mandin H, Brundin-Mather R. Reliability and validity of interviewers' judgements of medical school candidates. *Acad Med (Suppl)* 1996;**71**:40–2.

13 Eva KW. On the generality of specificity. *Med Educ* 2003;**37**:587–8**.**

14 Perkins DN, Salomon G. Are cognitive skills context-bound? *Educational Res* 1989;**18**:16–25.

15 Eva KW, Neville AJ, Norman GR. Exploring the aetiology of content specificity: factors influencing analogical transfer in problem solving. *Acad Med (Suppl)* 1998;**73**:1–5.

16 Ross L, Nisbett RE. *The Person and the Situation: Perspectives of Social Psychology*. New York: McGraw-Hill 1991.

17 Turnbull J, Danoff D, Norman GR. Content specificity and oral examinations. *Med Educ* 1996;**30**:56–9.

18 van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardised patients: state of the art. *Teach Learn Med* 1990;**2**:58–76.

19 Neville AJ, Cunnington JPW, Norman GR. Development of clinical reasoning exercises in a problem-based curriculum. *Acad Med (Suppl)* 1996;**71**:105–7.

20 Wood TJ, Cunnington JPW, Norman GR. Assessing the measurement properties of a clinical reasoning exercise. *Teach Learn Med* 2000;**12**:197–200.

21 Cunnington JPW, Neville AJ, Norman GR. The risks of thoroughness: reliability and validity of global ratings and checklists in an OSCE. *Adv Health Sci Educ* 1997;**1**:227–33.

22 Regehr G, Freeman R, Hodges B, Russell L. Assessing the generalisability of OSCE measures across content domains. *Acad Med* 1999;**74**:1320–2.

23 Kulatunga-Moruzi C, Norman GR. Validity of admissions measures in predicting performance outcomes: a comparison of those who were and were not accepted at McMaster. *Teach Learn Med* 2002;**14**:43–8.

Received 11 November 2002; editorial comments to authors 22 January 2003, 1 May 2003; accepted for publication 4 June 2003

APPENDIX

Station 1: Placebo (Ethical Decision Making)

Dr Cheung recommends homeopathic medicines to his patients. There is no scientific evidence or widely accepted theory to suggest that homeopathic medicines work, and Dr Cheung doesn't believe them to. He recommends homeopathic medicine to people with mild and non-specific symptoms such as fatigue, headaches and muscle aches, because he believes that it will do no harm, but will give them reassurance.

Consider the ethical problems that Dr Cheung's behaviour might pose. Discuss these issues with the interviewer.

Station 2: Aspartame (Critical Thinking)

A message that recently appeared on the Web warned readers of the dangers of aspartame (artificial sweetener – Nutrasweet, Equal) as a cause of an epidemic of multiple sclerosis (a progressive chronic disease of the nervous system) and systemic lupus (a multisystem auto-immune disease). The biological explanation provided was that, at body temperature, aspartame releases wood alcohol (methanol), which turns into formic acid, which 'is in the same class of drugs as cyanide and arsenic.' Formic acid, they argued, causes metabolic acidosis. Clinically, aspartame poisoning was argued to be a cause of joint pain, numbness, cramps, vertigo, headaches, depression, anxiety, slurred speech and blurred vision. The authors claimed that aspartame remains on the market because the food and drug industries have powerful lobbies in Congress. They quoted Dr Russell Blaylock, who said, 'The ingredients stimulate the neurons of the brain to death, causing brain damage of varying degrees.'

Critique this message, in terms of the strength of the arguments presented and their logical consistency. Your critique might include an indication of the issues that you would like to delve into further before assessing the validity of these claims.

Station 3: Air Travel (Communication Skills)

Your company needs both you and a co-worker (Sara, a colleague from another branch of the company) to attend a critical business meeting in San Diego. You have just arrived to drive Sara to the airport.

Sara is in the room.

Station 4: Deterrent Fees (Knowledge of the Health Care System)

Recently, the Prime Minister of Canada raised the issue of deterrent fees (a small charge, say $10, which everyone who initiates a visit to a health professional would have to pay at the first contact) as a way to control health care costs. The assumption is that this will deter people from visiting their doctor for unnecessary reasons.

Consider the broad implications of this policy for health and health care costs. For example, do you think the approach will save health care costs? At what expense? Discuss this issue with the interviewer.

Station 5: Standard Interview 1

Why do you want to be a physician? Discuss this question with the interviewer.

Station 6: Circumcision (Ethical Decision Making)

The Canadian Pediatric Association has recommended that circumcisions 'not be routinely performed'. They base this recommendation on their determination that 'the benefits have not been shown to clearly outweigh the risks and costs'. Doctors have no obligation to refer for, or provide, a circumcision, but many do, even when they are clearly not medically necessary. Ontario Health Insurance Plan (OHIP) no longer pays for unnecessary circumcisions.

Consider the ethical problems that exist in this case. Discuss these issues with the interviewer.

Station 7: Class Size (Critical Thinking)

Universities are commonly faced with the complicated task of balancing the educational needs of their

students and the cost required to provide learning resources to a large number of individuals. As a result of this tension, there has been much debate regarding the optimal size of classes. One side argues that smaller classes provide a more educationally effective setting for students, while others argue that it makes no difference, so larger classes should be used to minimise the number of instructors required.

Discuss your opinion on this issue with the examiner.

Station 8: Parking Garage (Communication Skills)

The parking garage at your place of work has assigned parking spots. On leaving your spot, you are observed by the garage attendant as you back into a neighbouring car, a BMW, knocking out its left front headlight and denting the left front fender. The garage attendant gives you the name and office number of the owner of the neighbouring car, telling you that he is calling ahead to the car owner, Tim. The garage attendant tells you that Tim is expecting your visit.

Enter Tim's office.

Station 9: Preferential Admission (Knowledge of the Health Care System)

Due to the shortage of physicians in rural communities such as those in Northern Ontario, it has been suggested that medical programmes preferentially admit students who are willing to commit to a 2- or 3-year tenure in an under-serviced area upon graduation.

Consider the broad implications of this policy for health and health care costs. For example, do you think the approach will be effective? At what expense? Discuss this issue with the interviewer.

Station 10: Standard Interview 2

What experiences have you had (and what insights have you gained from these experiences) that lead you to believe you would be a good physician?

Discuss this question with the interviewer.