Predictive validity of the multiple mini-interview for selecting medical trainees

Kevin W Eva,[1] Harold I Reiter,[2] Kien Trinh,[3] Parveen Wasi,[4] Jack Rosenfeld[5] & Geoffrey R Norman[1]

INTRODUCTION In this paper we report on further tests of the validity of the multiple mini-interview (MMI) selection process, comparing MMI scores with those achieved on a national high-stakes clinical skills examination. We also continue to explore the stability of candidate performance and the extent to which so-called 'cognitive' and 'non-cognitive' qualities should be deemed independent of one another.

METHODS To examine predictive validity, MMI data were matched with licensing examination data for both undergraduate ($n = 34$) and postgraduate ($n = 22$) samples of participants. To assess the stability of candidate performance, reliability coefficients were generated for eight distinct samples. Finally, correlations were calculated between 'cognitive' and 'non-cognitive' measures of ability collected in the admissions procedure, on graduation from medical school and 18 months into postgraduate training.

RESULTS The median reliability of eight administrations of the MMI in various cohorts was 0.73 when 12 10-minute stations were used with one examiner per station. The correlation between performance on the MMI and number of stations passed on an objective structured clinical examination-based licensing examination was $r = 0.43$ ($P < 0.05$) in a postgraduate sample and $r = 0.35$ ($P < 0.05$) in an undergraduate sample of subjects who sat the MMI 5 years prior to sitting the licensing examination. The correlation between 'cognitive' and 'non-cognitive' assessment instruments increased with time in training (i.e. as the focus of the assessments became more tailored to the clinical practice of medicine).

DISCUSSION Further evidence for the validity of the MMI approach to making admissions decisions has been provided. More generally, the reported findings cast further doubt on the extent to which performance can be captured with trait-based models of ability. Finally, although a complementary predictive relationship has consistently been observed between grade point average and MMI results, the extent to which cognitive and non-cognitive qualities are distinct appears to depend on the scope of practice within which the two classes of qualities are assessed.

Medical Education 2009: 43: 767–775
doi:10.1111/j.1365-2923.2009.03407.x

[1]Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada
[2]Department of Oncology, McMaster University, Hamilton, Ontario, Canada
[3]Department of Family Medicine, McMaster University, Hamilton, Ontario, Canada
[4] Department of Medicine, McMaster University, Hamilton, Ontario, Canada
[5]Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada

Correspondence: Kevin W Eva, Department of Clinical Epidemiology and Biostatistics, Program for Educational Research and Development, MDCL 3522, McMaster University, Hamilton, Ontario L8N 3Z5, Canada. Tel: 00 1 905 525 9140 (ext 27241); Fax: 00 1 905 572 7099; E-mail: evakw@mcmaster.ca

INTRODUCTION

The literature on medical school admissions consistently draws a distinction between cognitive and non-cognitive abilities.[1] Broadly defined, the former refers to intellectual prowess, typically measured by grade point average (GPA) or performance on standardised tests of knowledge, such as the Medical College Admission Test (MCAT)[2] in North America or the Graduate Australian Medical School Admission Test (GAMSAT) in Australia and the UK.[3] The latter quality, non-cognitive aptitude, is typically used to encapsulate all the other qualities that might be desired in an applicant. The list of non-cognitive characteristics is likely to be endless – Albanese *et al.* have used the literature to identify over 80 such qualities[4] – but it includes the abilities to relate to others on a personal level, to make sound clinical and professional judgements, and to carry out ethical decision making. Although there is some variation in opinion with respect to the relative weight that should be assigned to cognitive and non-cognitive measures of potential at medical school admissions, there is widespread agreement that it is desirable to broaden the scope of assessment beyond academic achievement.[5]

At the same time, dissociating aspects of competence in this way and debating their relative value can result in caricatures becoming perceived as the norm. One such caricature is that of the bookworm, which portrays an individual who performs exceptionally well in school, but who has difficulty conversing with others or appears to have little practical knowledge. Another is the butterfly, which describes an individual who has excellent interpersonal qualities, but who is generally challenged when attempting to reason through intellectual dilemmas. Common awareness of individuals who fit those descriptions can make it seem as though cognitive and non-cognitive qualities are negatively correlated, yet there is an argument to be made that success on one side of the ledger (e.g. in cognitive aptitude) enables success on the other side (e.g. in non-cognitive aptitude) and vice versa. Recent evidence has suggested that GPA may, in fact, be a better indicator of work ethic or dedication (i.e. of non-cognitive qualities) than it is of intelligence.[6] At the same time, studies in medicine and elsewhere suggest that knowledge, in large part, is the foundation on which other competencies are built.[7,8] Even at an intuitive level it seems unlikely that an individual could ever communicate well with patients if he or she does not have knowledge to communicate.

Until recently, however, it has been difficult to grapple with issues of how various qualities interrelate because it was unclear how much one should trust the measurement instruments that were commonly used to assess non-cognitive qualities. The ever-present impact that context has on performance[9] makes it difficult to know if poor correlations between domains indicate independence of the underlying constructs or result simply from doomed attempts to correlate anything with the noise produced by weak measurement tools.

With the production of recent evidence supporting the validity of multiple mini-interview (MMI) procedures[10] and the objective structured clinical examination (OSCE)-based national licensing examination used in Canada,[11] to name two examples, we are now able to begin a more informed exploration of how these various qualities interrelate. To this end, the purpose of this paper is three-fold:

1. to continue exploration of the MMI format's potential to discriminate between applicants in a reliable manner by studying an arguably more homogeneous cohort of participants (postgraduate trainees) than has been reported on in the past;
2. to further test the ability of the MMI process to validly predict the non-cognitive competencies of practising doctors by comparing scores achieved on the MMI with those achieved on Part II of the Medical Council of Canada Qualifying Examination (MCCQE Part II), a national, high-stakes clinical skills examination required for licensure, and
3. to further consider the relationship between measures of 'cognitive' and 'non-cognitive' competencies in medical settings.

In considering these goals it should be noted that the MMI is a process rather than a test. At McMaster University, our stakeholders desired that the MMI be heavily weighted towards ethical decision making and that some attention be paid to communication skills and collaborative ability.[12] It is for this reason that we believe we have seen a significant correlation between our MMI and the ethics component of Part I of the MCCQE,[10] and why we consider Part II an important criterion against which to compare both undergraduate and postgraduate students. The MMIs developed in other places may be aimed at different characteristics[13] and hence may anticipate different results.

ISSUE 1. THE STABILITY OF PERFORMANCE

Methods

The goal in most measurement contexts is to measure some 'characteristic' of the individual (i.e. how much of a competency he or she possesses). Thinking of these competencies as traits, however, may be fallacious, given that previous work on the MMI and in other measurement contexts has suggested that context specificity rules the day (i.e. performance in one setting tends to be poorly predictive of performance in other settings).[9] To further test the stability of performance (and to examine the dominant sources of error), generalisability analyses were performed on a sample of postgraduate trainees. These individuals were accepted to (and completed) undergraduate medical training, after which they chose (and were selected to) postgraduate training in a medical specialty. As a result, they were expected to represent a more homogeneous population than the broad pool of applicants to an undergraduate medical programme.

In July 2005, a sample of 29 second- and third-year residents were recruited from the internal medicine (n = 27) and radiation oncology (n = 2) programmes at McMaster University to participate in a nine-station MMI. Subjects were compensated financially for their participation. Their average age was 28.7 years (standard deviation 3.8 years) and the group was evenly split in terms of gender (13 women, 14 men; two participants did not specify their gender on the demographics form). The MMI was blueprinted as described above (i.e. with a heavy emphasis on ethical decision making and a non-clinical focus) and was run in accordance with the protocol developed by Eva *et al.*[14] Participants were assigned to one of three circuits that ran sequentially (two on a single day, one a week later). Each circuit was composed of nine 10-minute stations (the same stations were used each day), each of which was staffed by one examiner. Participant performance was rated after each station using a scoresheet that contained four items with accompanying 7-point global rating scales. A total of 18 examiners participated across the three circuits (nine per day). The examiners were recruited from the wider health sciences faculty and included a midwife, a pathologist, a radiation therapist, and staff doctors and senior residents drawn from a variety of residency programmes.

Generalisability theory was used to assess the stability of performance and the reliability of the scores assigned. To ensure the consistency of the findings, we also examined (and briefly report on) the reliability found in the first eight administrations of the MMI run at McMaster University. Although the methodology and strategy adopted in each instance were consistent with the methodology described above, the number of stations used varied from three to 12. Decision studies were used to equate each administration to the 12-station, one-examiner-per-station protocol now being used by the undergraduate MD programme at McMaster University.

RESULTS

Although the residents included in this study were a unique and arguably more homogeneous sample, given their postgraduate and specialist status relative to the samples included in previous reports,[14–16] reliabilities were found to be much the same. The inter-station reliability of a single interview (station) was found to be G = 0.24, whereas the reliability of the total score generated by averaging across all nine stations equalled 0.76. A decision study suggested that a 12-station MMI in this context would yield reliability of 0.80. The variance components and formulae used to generate these reliability coefficients are illustrated in Table 1. The resident × station (nested within circuit) interaction was the greatest contributor to the error variance, but it should be noted that rater and station are completely confounded in this study design, unlike in some prior designs.[17]

These results are highly consistent with those of a number of other studies of the MMI. The reliability of any one station is consistently low, typically < 0.25, whereas the reliability of the average performance score across 12 stations has been found to have a median reliability of 0.73. Table 2 illustrates these findings by reporting the reliability of a 12-station MMI as determined using D studies performed on the first eight administrations of the MMI completed at McMaster University.

ISSUE 2. PREDICTIVE VALIDITY OF THE MMI

Methods

Previous work has shown a complementary relationship between the predictive capacity of GPA and the MMI. GPA has predominantly predicted knowledge-based measures like intramural performance on tests of medical knowledge[18] and on the core clerkship

Table 1 Variance components and generalisability coefficients for a multiple mini-interview run with postgraduate second- and third-year residents (n = 29)

Source of variance	Degrees of freedom	Variance component	Percent of total variance
Circuit	2	0	0
Resident:Circuit	26	0.82	19.71
Station	8	0.40	9.62
Item	3	0.01	0.24
Circuit*Station	16	0.68	16.35
Circuit*Item	6	0	0
Resident*Station:Circuit	208	1.84	44.23
Resident*Item:Circuit	78	0.02	0.48
Station*Item	24	0.02	0.48
Circuit*Station*Item	48	0.01	0.24
Residual error (rsi:c)	624	0.36	8.65

$G(1) = [v(r:c)]/[v(r:c) + v(ri:c) + v(rs:c) + v(s) + v(si) + v(rsi:c)] = 0.24$

$G(9) = [v(r:c)]/[v(r:c) + v(ri:c) + v(rs:c)/9 + v(s)/9 + v(si)/9 + v(rsi:c)/9] = 0.76$

$G(12) = [v(r:c)]/[v(r:c) + v(ri:c) + v(rs:c)/12 + v(s)/12 + v(si)/12 + v(rsi:c)/12] = 0.80$

Table 2 Generalisability and administration details of the multiple mini-interview (MMI) in the first eight administrations

Year	Group	*n*	Stations used	Examiners per station	Reliability observed	Reliability (calculated for 12 stations with one examiner per station)
2002	MD programme[14]	117	10	1	0.65	0.69
2003	MD programme[15]	57	9	2	0.78	0.74
2003	Rehabilitation science	57	3	2	0.51	0.77
2004	MD programme*	379	12	1	0.70	0.70
2004	Rehabilitation science*	380	7	1	0.70	0.79
2005	MD programme*	696	12	1	0.69	0.69
2005	Internal medicine	29	9	1	0.69	0.76
2005	Emergency medicine*	30	10	2	0.75	0.72

* In these administrations the MMI was used to make the selection decision rather than being mounted for research purposes only

rotation sub-scores of the MCCQE Part I.[10] By contrast, the MMI has predominantly predicted performance-oriented measures, such as clinical clerkship ratings, and ethics-based measures, like the *c*ultural, *c*ommunication, *l*egal, *e*thical and *o*rganisational competencies (C2LEO) component of the MCCQE Part I.[10] In this paper we report on the relationship between the MMI and Part II of the MCCQE.

The two-part MCCQE is a national examination all MDs must pass to become licentiates of the Medical Council of Canada (MCC). The licentiate is one of the prerequisites to licensure in Canada.[19] The MCCQE Part I is a computer-based test primarily aimed at assessing medical knowledge and clinical decision-making skills that is typically administered within 2 weeks of a candidate's completion of undergraduate medical training. The MCCQE Part II requires examinees to pass an OSCE during which they are observed interacting with a series of standardised patients by doctor-examiners.[20] To be eligible to take Part II, examinees must have passed Part I and have a minimum of 1 year of clinical postgraduate training. Although succeeding on Part II requires medical knowledge, greater emphasis is placed on clinical performance and communication skills than in Part I. Performance reports include scores associated with the domains of Data Acquisition (performing a proper history and physical examination), Problem Solving (synthesising the information provided appropriately), Patient Interaction (communication skills) and the C2LEO. A total score is assigned using examiner-completed checklists. A modified borderline groups method[21] is used to set a cut-score for each station and, to pass, examinees must achieve a mark greater than the cut-score for a specified percentage of stations. As the MCCQE Part II is intended to reflect more of the non-cognitive competencies of doctor performance than is Part I, we anticipated that the MMI would provide the best prediction of Part II scores.

To examine this prediction we correlated scores achieved on the MMI in the postgraduate study described in the preceding section with those achieved on Part II. These participants sat Part II either 9 months prior to sitting the MMI or 3 months afterwards, depending on their year of training. The datasets were combined by a third party such that the MCC did not receive local data and local investigators were unable to link MCC data to participant identities. As a result of this process, matches were achieved for only 22 of the 29 participants. As this is a relatively small sample, we attempted to replicate the study with a second sample of participants that better represented the time lag between undergoing undergraduate admissions processes and completing the Part II examination.

In the winter of 2002, 117 applicants to McMaster University's undergraduate MD programme participated in an MMI in addition to undertaking four other admissions procedures: upon application, these individuals submitted a transcript (to enable calculation of their GPA) and an autobiographical submission (ABS; a written response to a series of questions pertaining to the applicant's personal history, suitability for medical training and goals for future practice). These two measures were used to select applicants for interview. At that point, candidates completed a traditional, panel-based interview with three interviewers, were observed interacting with other candidates during a simulated tutorial exercise (also rated by three assessors), and completed an MMI consisting of 10 stations, each of which was scored by one interviewer.[14] Forty-five of the 117 candidates gained entry into the programme and have subsequently graduated. Previously reported analyses showed these 45 to be representative of the larger sample.[10] The reliability of the MMI and the relationship between the admissions measures and written performance indicators (both intramurally and on the MCCQE Part I) have been reported previously.[10,18] Since that last report we have been able to use the same matching process as described for the postgraduate sample to compare admissions scores with Part II results for 34 participants who sat the MCCQE Part II in October 2007.

RESULTS

As described above, the key result on the MCCQE Part II is the percentage of stations passed as this determines whether or not a candidate becomes a licentiate of the Medical Council of Canada. In the postgraduate sample the MMI was statistically predictive of the percentage of stations candidates passed on the MCCQE Part II examination ($r = 0.43$, $P < 0.05$) and trended towards being statistically predictive of total score ($r = 0.36$, $P < 0.1$). Table 3 illustrates that this finding appears to be driven by the ability of the MMI to predict Patient Interaction and C2LEO scores, whereas the MMI was not related to scores assigned to candidates' Data Acquisition or Problem Solving performances.

One advantage of the undergraduate sample over the postgraduate sample is that data were also available for other admissions tools (GPA, ABS, simulated tutorial and a personal interview) to enable a comparison of the relative predictive value of the MMI. Conversely, in the undergraduate sample, a 5.5-year time lag occurred between the tests, making it likely

Table 3 Correlation between multiple mini-interview scores and Medical Council of Canada Qualifying Examination Part II scores in the postgraduate sample

Part II score	*r*	*p*
Number of stations passed	0.43	< 0.05
Total score	0.36	< 0.10
Patient interaction sub-score	0.65	< 0.01
C2LEO sub-score	0.44	< 0.05
Data acquisition sub-score	0.09	NS
Problem solving sub-score	0.05	NS

C2LEO = communication, cultural, *l*egal, ethical and organisational aspects of medicine; NS = not significant

Table 4 Correlation between admissions tools and Medical Council of Canada Qualifying Examination Part II scores for the undergraduate sample

Admissions tool	$r_{\text{(number of stations passed)}}$	p	$r_{\text{(total score)}}$	p
Autobiographical sketch	0.14	NS	0.03	NS
Grade point average	− 0.03	NS	0.04	NS
Multiple mini-interview	0.35	< 0.05	0.19	NS
Personal interview	0.13	NS	0.03	NS
Simulated tutorial	0.07	NS	− 0.10	NS

NS = not significant

that all correlations would be reduced. Nonetheless, the MMI was found to be the only statistically significant predictor of the percentage of stations that candidates passed on the MCCQE Part II (Table 4). Although none of the tools significantly predicted total score on the Part II, the MMI revealed the highest correlation with the total score.

ISSUE 3. THE RELATIONSHIP BETWEEN MEASURES OF 'COGNITIVE' AND 'NON-COGNITIVE' COMPETENCIES

Methods

At three time-points, scores generally intended to reflect 'cognitive' competencies were correlated with scores generally intended to reflect 'non-cognitive' competencies. The first time-point was the time of application to medical school, correlating MMI score with GPA for the winter 2002 group of 117 applicants described above. The second time-point was the conclusion of medical school training, correlating the core rotation sub-scores on the MCCQE Part I with the C2LEO scores on the same examination. The third time-point occurred during residency training, correlating the predominantly cognitive domains of Part II (Data Gathering and Problem Solving) and the predominantly non-cognitive domains of the same examination (Patient Interaction and C2LEO).

RESULTS

At the time of application in winter of 2002, the correlation between GPA and MMI score, for 117 applicants, was $r = -0.23$ ($P < 0.01$). Within the MCCQE Part I examination the correlation between core rotation sub-scores and the C2LEO scores for the 45 of those applicants who were admitted was $r = 0.17$ ($P > 0.10$). The correlation between the predominantly cognitive domains of Part II (Data Gathering and Problem Solving) and the predominantly non-cognitive domains of the same examination (Patient Interaction and C2LEO) for the 34 matriculants with matched data was $r = 0.27$ ($P < 0.10$). The same comparison for the 22 residents with available data who sat the MMI was $r = 0.43$ ($P < 0.05$).

DISCUSSION

The reliability and validity of any assessment strategy is dependent on the context within which the strategy is applied and the content of the assessment. Through previous research efforts we have reported that the reliability of a 12-station OSCE-type MMI procedure, blueprinted to assess ethics, communication skills and collaborative ability predominantly, achieves acceptable levels of reliability and validity in the context of undergraduate medical school admissions.[10,14] This is the first paper to report on the reliability of the process at the postgraduate level, an important finding as this sample could intuitively be thought of as more homogeneous (i.e. more difficult to differentiate) because almost all participants were pre-selected into a particular specialty (internal medicine). As shown in Table 2, similar results were found regardless of whether the MMI was used to select to undergraduate MD, rehabilitation science or postgraduate training programmes and whether or not the process was used to determine who gained entry into the health professional programme rather than being used simply for research or feasibility testing. Other institutions have begun to report similar findings based upon their local experiences.[15,16]

This paper also reports further evidence of the validity of the MMI process. In two distinct samples the MMI was found to be the most predictive measure of performance on Part II of the MCCQE. The magnitude of the correlations was more impressive in the postgraduate than in the undergraduate sample, but it should be remembered that the time lag between sitting the MMI and sitting the MCCQE Part II was over 5 years for the latter group, and that this included the period between application and matriculation, 3 years of undergraduate training and 16 months of postgraduate training. Any correlation in this context of countless confounding variables stands in contrast to the small and non-significant correlations found between the MCCQE Part II and the other admissions instruments included in this and other studies.[22]

This study is clearly limited by the sample size of participants available for analysis. In an effort to overcome this limitation, we focused our efforts on replication and we derive confidence from the consistency of the findings across independent cohorts of participants. Ideally, actual practice outcomes would be used to determine the extent to which any of the measures collected (admissions, intramural or licensing examination performance) are predictive of clinical performance. Unfortunately, the admissions process under study has not been in existence long enough to allow the consideration of such outcomes, but we anticipate performing such follow-up studies in the future.

Even without long-term follow-up, however, the findings reported take on added importance, given the recent report by Tamblyn *et al.*,[11] which revealed a relationship between performance on the Part II examination and complaints to regulatory authorities up to 12 years post-graduation. As a multitude of variables are likely to impact upon an individual's growth throughout his or her career, a chain of evidence is probably the best option in terms of optimising the various admissions and selection decisions that are made in the course of one's medical training and career. To this point we have seen evidence suggesting that incoming GPA predicts knowledge-oriented intramural measures,[10,18] that those measures are predictive of licensing examination performance,[23] and that licensing examination performance can be predictive of outcomes as real as patient mortality rates 10 years into practice.[24] In terms of non-cognitive attributes, we can now report that the MMI predicted intramural clinical performance measures and performance-based licensing examination scores; in addition, the work of Papadakis *et al.*[25] and Tamblyn *et al.*[11] suggests that these outcomes predict professional lapses in actual practice. Of course, we cannot guarantee that just because *a* predicts *b* and *b* predicts *c*, that *a* will predict *c* and, indeed, the correlations between these links are moderate. Still, they are the largest reported to date and in view of the complexity of the context and long delays, they may be all we can expect.

The broader goal of this programme of research, however, has been to refine understanding of the challenges inherent in selection. Although it is certainly the case that raters' opinions of candidates vary, this programme of research has suggested that more is gained psychometrically by collecting ratings from multiple people spread over multiple interviews than by simply increasing the number of raters within an interview.[17] This may not be a universal phenomenon, but determining when it is the case and when it is not is an area requiring further research and the bottom line, pragmatically speaking, is that, regardless of the cause, overcoming the problem of measurement error requires results to be averaged across many observations of performance in admissions contexts, just as we routinely average across many observations when conducting knowledge tests.

With that as background, this line of exploration has changed our perceptions of the relationship between personal and intellectual tendencies. We have heard concerns that using cognitive measures like GPA or MCAT scores to make selection decisions will result in a class of asocial bookworms. We have, nevertheless, argued that intelligence and interpersonal tendencies are not mutually exclusive.[26] The original argument was derived from the lack of relationship between cognitive and non-cognitive admissions measures. Further indirect evidence for the independence of cognitive and non-cognitive tendencies can be derived from the finding of complementarity with respect to outcomes predicted by GPAs and MMI scores.[10] The correlations reported here, however, and others in the literature would now lead us to believe that, if anything, as we delve deeper into a particular domain of study or practice, it becomes more difficult to claim independence between the cognitive and non-cognitive domains. Counter to concerns about asocial bookworms it would appear that, if anything, the relationship is positive rather than negative. Although the correlation between GPA and MMI outcomes was found to be slightly negative, the relationships between cognitive and non-cognitive subdomains of the licensing examinations reported here ranged between $r = 0.17$ and $r = 0.43$ and correlations generally increased with trainees' seniority. The literature has shown that one of the better predictors of clinical skills (including interpersonal tendencies and communication skills) is performance on the verbal reasoning portion of the MCAT,[27] a decidedly cognitive measure. Further, the relationship between undergraduate GPA and doctor disciplinary action reported by Papadakis *et al.* was small, but statistically significant in two studies, both of which showed trends indicating that doctors with professional behaviour difficulties tended to have lower GPAs.[25,28]

There are many possible reasons for the slight to moderate correlations found between cognitive and non-cognitive competencies that cannot yet be fully disentangled. It would seem, however, that within a domain like medicine (as opposed to across the broader scope of pre-medical measurement instruments), the differing competencies may not be completely independent. Delivering bad news well to a cancer patient requires an acute knowledge of prognosis, as well as a proclivity to communicate well. Clinical decision making requires an understanding not only of the nature of disease, but also of its implications on the unique sensibilities of the patient in question. This perspective has implications in terms of admissions philosophy and procedures. The desired end-product is neither a set of bookworms nor a set of butterflies, but a cohort of well-rounded professionals who combine the best characteristics of both. Fortunately, the typical applicant pool to most medical schools allows for such elitism, for there will very likely be a significant number of applicants who meet these multiple expectations.

Contributors: all authors contributed to the study design. KWE performed the data analyses and drafted the manuscript. All other authors offered critical revisions and all authors approved the final manuscript for publication.
Acknowledgements: the authors would like to thank Julijana Malinovic for her assistance with data collection, Robert Lee, Steve Slade and Tim Wood for their assistance with data merging, and Mary Lou Schmuck for her assistance with data analysis.
Funding: this project was funded (in part) by a National Board of Medical Examiners® (NBME®) Edward J Stemmler MD Medical Education Research Fund grant. The project does not necessarily reflect NBME policy and NBME support does not represent official endorsement.
Conflicts of interest: HIR and JR are developing a commercial enterprise to market their expertise with admissions strategies, including the multiple mini-interview. No other authors have competing interests.
Ethical approval: all of the research reported in this paper was approved by the Hamilton Health Sciences Research Ethics Board.

REFERENCES

1 Benbassat J, Baumal R. Uncertainties in the selection of applicants for medical school. *Adv Health Sci Educ Theory Pract* 2007;**12** (4):509–21.
2 Association of American Medical Colleges. Medical College Admissions Test. http://www.aamc.org/students/mcat/. [Accessed 12 June 2009.]
3 Australian Council for Educational Research. Graduate Australian Medical School Admissions Test. http://www.gamsat.edu.au/. [Accessed 12 June 2009.]
4 Albanese MA, Snow MH, Skochelak SE, Huggett KN, Farrell PM. Assessing personal qualities in medical school admissions. *Acad Med* 2003;**78**:313–21.
5 Institute for International Medical Education. Global minimum essential requirements in medical education. Report of the Core Committee. *Med Teach* 2002;**24**:130–5.
6 Duckworth AL, Seligman MEP. Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychol Sci* 2005;**16** (12):939–44.
7 Norman GR, Tugwell P, Feightner JW, Muzzin LJ, Jacoby LL. Knowledge and clinical problem solving. *Med Educ* 1985;**19** (5):344–56.
8 Patel VL, Glaser R, Arocha JF. Cognition and expertise: acquisition of medical competence. *Clin Invest Med* 2000;**23** (4):256–60.
9 Eva KW. On the generality of specificity. *Med Educ* 2003;**37** (7):587–8.
10 Reiter HI, Eva KW, Rosenfeld J, Norman GR. Multiple mini-interviews predict clerkship and licensing examination performance. *Med Educ* 2007;**41** (4):378–84.
11 Tamblyn R, Abrahamowicz M, Dauphinee D *et al.* Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *JAMA* 2007;**298** (9):993–1001.
12 Reiter HI, Eva KW. Reflecting the relative values of community, faculty, and students in the admissions tools of medical school. *Teach Learn Med* 2005;**17**:4–8.
13 Roberts C, Walton M, Rothnie I, Crossley J, Lyon P, Kumar K, Tiller D. Factors affecting the utility of the multiple mini-interview in selecting candidates for graduate entry medical school. *Med Educ* 2008;**42**:396–404.
14 Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Med Educ* 2004;**38** (3):314–26.
15 Lemay JF, Lockyer JM, Collin VT, Brownell AK. Assessment of non-cognitive traits through the admissions multiple mini-interview. *Med Educ* 2007;**41** (6):573–9.
16 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** (3):234–41.
17 Eva KW, Reiter HI, Rosenfeld J, Norman GR. The relationship between interviewers' characteristics and ratings assigned during a multiple mini-interview. *Acad Med* 2004;**79** (6):602–9.
18 Eva KW, Reiter HI, Rosenfeld J, Norman GR. The ability of the multiple mini-interview to predict preclerkship performance in medical school. *Acad Med* 2004;**10** (Suppl):40–2.
19 Medical Council of Canada. About the examinations. http://www.mcc.ca/en/exams/. [Accessed 12 June 2009.]
20 Reznick RK, Blackmore D, Dauphinee WD, Rothman AI, Smee S. Large-scale high-stakes testing with an OSCE: report from the Medical Council of Canada. *Acad Med* 1996;**1** (Suppl):19–21.
21 Dauphinee WD, Blackmore DE, Smee S, Rothman AI, Reznick R. Using the judgements of physician examiners in setting the standards for a national multicentre

high-stakes OSCE. *Adv Health Sci Educ Theory Pract* 1997;**2** (3):201–11.
22 Kulatunga-Moruzi C, Norman GR. Validity of admissions measures in predicting performance outcomes: the contribution of cognitive and non-cognitive dimensions. *Teach Learn Med* 2002;**14**:34–42.
23 Blake JM, Norman GR, Smith EKM. Report card from McMaster: student evaluation at a problem-based medical school. *Lancet* 1995;**345**:899–902.
24 Norcini JJ, Lipner RS, Kimball HR. Certifying examination performance and patient outcomes following acute myocardial infarction. *Med Educ* 2002;**36**:853–9.
25 Papadakis MA, Teherani A, Banach MA, Knettler TR, Rattner SL, Stern DT, Veloski JJ, Hodgson CS. Disciplinary action by medical boards and prior behaviour in medical school. *N Engl J Med* 2005;**25**:2673–82.
26 Eva KW, Reiter HI. Where judgement fails: pitfalls in the selection process for medical personnel. *Adv Health Sci Educ Theory Pract* 2004;**9** (2):161–74.
27 Donnon T, Paolucci EO, Violato C. The predictive validity of the MCAT for medical school performance and medical board licensing examinations: a meta-analysis of the published research. *Acad Med* 2007;**82** (1):100–6.
28 Papadakis MA, Hodgson CS, Teherani A, Kohatsu ND. Unprofessional behaviour in medical school is associated with subsequent disciplinary action by a state medical board. *Acad Med* 2004;**79** (3):244–9.

Received 26 November 2008; comments to authors 12 January 2009; accepted for publication 30 March 2009