

Multiple mini-interviews predict clerkship and licensing examination performance

HAROLD I REITER,¹ KEVIN W EVA,² JACK ROSENFELD³ & GEOFFREY R NORMAN²

OBJECTIVE The Multiple Mini-Interview (MMI) has previously been shown to have a positive correlation with early medical school performance. Data have matured to allow comparison with clerkship evaluations and national licensing examinations.

METHODS Of 117 applicants to the Michael G DeGroot School of Medicine at McMaster University who had scores on the MMI, traditional non-cognitive measures, and undergraduate grade point average (uGPA), 45 were admitted and followed through clerkship evaluations and Part I of the Medical Council of Canada Qualifying Examination (MCCQE). Clerkship evaluations consisted of clerkship summary ratings, a clerkship objective structured clinical examination (OSCE), and progress test score (a 180-item, multiple-choice test). The MCCQE includes subsections relevant to medical specialties and relevant to broader legal and ethical issues (Population Health and the Considerations of the Legal, Ethical and Organizational Aspects of Medicine [CLEO/PHELO]).

RESULTS In-programme, MMI was the best predictor of OSCE performance, clerkship encounter cards, and clerkship performance ratings. On the MCCQE Part I, MMI significantly predicted CLEO/PHELO scores and clinical decision-making (CDM) scores. None of these assessments were predicted by other non-cognitive admissions measures or uGPA. Only uGPA predicted progress test scores and the MCQ-based specialty-specific subsections of the MCCQE Part I.

¹Department of Oncology, McMaster University, Hamilton, Ontario, Canada

²Department of Clinical Epidemiology and Biostatistics, Program for Educational Research and Development, McMaster University, Hamilton, Ontario, Canada

³Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada

Correspondence: Kevin W Eva, MDCL 3522, McMaster University, 1200 Main Street West, Hamilton, Ontario L8N 3Z5, Canada.
Tel: 00 1 905 525 9140 ext. 27241; Fax: 00 1 905 572 7099;
E-mail: evakw@mcmaster.ca

DISCUSSION The MMI complements pre-admission cognitive measures to predict performance outcomes during clerkship and on the Canadian national licensing examination.

KEYWORDS clinical clerkship/ *standards; clinical competence/ *standards; *licensure, medical; Ontario; school admission criteria; schools, medical.

Medical Education 2007; **41**: 378–384

doi:10.1111/j.1365-2929.2007.02709.x

INTRODUCTION

Pre-admission measures for medical school are grouped roughly into those that predominantly represent cognitive and those that predominantly represent professional (i.e. non-cognitive) domains. Within the former, undergraduate grade point average (uGPA) and aptitude tests like the Medical College Admissions Test (MCAT) have repeatedly demonstrated reliability and predictive validity.^{1,2} Observed validity coefficients for uGPA with in-school and licensure measures have been in the order of 0.26–0.40; those for MCAT in the order of 0.32–0.61.² The same cannot be said regarding measures of professional domains, such as personal interviews, autobiographical assessments, reference letters^{3,4} or group interviews/simulated tutorials,⁵ for which predictive validity correlations rarely rise above 0.10. Context specificity is a significant limitation, restricting the reliability, and hence usefulness, of historical measures of non-cognitive tendencies as they rely on a single occasion of measurement. These concerns led to the development of the Multiple Mini-Interview (MMI) at McMaster University in 2001.

The MMI logistically resembles an admissions objective structured clinical examination (OSCE) with multiple short-interview stations. Unlike an OSCE,

Overview

What is already known on this subject

The Multiple Mini-Interview (MMI) is a feasible, acceptable and reliable admissions protocol that is predictive of pre-clerkship OSCE performance.

What this study adds

This study provides further support for the MMI as a measure of non-cognitive (i.e. personal) qualities in medical school applicants. Relative to other admissions tools, the MMI was the best predictor of intramural clinical performance ratings and ethical/clinical decision-making scores on the Canadian national licensing examination.

Suggestions for further research

The relationship between MMI scores and scores on the OSCE-based component of the Canadian licensing examination has yet to be considered, as does the adequacy of the MMI for making postgraduate admissions decisions.

however, the MMI focuses exclusively on non-cognitive domains such as moral reasoning, communication skills and the ability to collaborate. The stations are much more open-ended than those in an OSCE; interviewer–applicant interaction is supported by station-related background and theory provided to the interviewer. By design, none of the stations anticipate prior health care knowledge on the part of the applicant. By obtaining multiple biopsies of non-cognitive abilities, the MMI has demonstrated much higher reliability than standard interview methods, with a published overall test generalisability of 0.65,⁶ and with results on dozens of subsequent MMI administrations carried out at multiple schools since then averaging around 0.75. However, reliability alone is insufficient to demonstrate the usefulness of the MMI. Demonstration of predictive validity by uGPA and MCAT has been the hallmark of their usefulness; the same expectation should be set for all admissions instruments.

The MMI has already demonstrated some evidence of validity. In 2004, Eva *et al.* reported that the MMI

correlated better with pre-clerkship OSCEs than any other admissions instrument used at McMaster.⁷ Grade point average was most predictive of progress test (periodic multiple-choice test) performance. In-school predictive criterion measures are useful, but they represent relatively short-term outcomes and may therefore be less important than longer-term measures that are more closely related with post-training clinical performance. Such measures include clinical performance during clerkship⁸ and performance on national licensing examinations.⁵ To date, various measures of non-cognitive qualities have been shown to be non-predictive of the latter performance measures.⁵

Since the spring of 2005, a cohort of 45 medical school applicants who were admitted to medical school and took part in the first large-scale MMI have graduated from McMaster University and have written the first part of the Medical Council of Canada's Qualifying Examination (the MCCQE Part I). The examination consists of a multiple-choice question (MCQ) component taken in the morning and a patient-based, short-answer clinical decision-making (CDM) 'Key Features' component^{9,10} taken in the afternoon. Both morning and afternoon components are computer-based and cover basic medical knowledge in the following areas: medicine; surgery; obstetrics and gynaecology; psychiatry; paediatrics, and a sixth area entitled Population Health and the Considerations of the Legal, Ethical and Organisational Aspects of Medicine (CLEO/PHELO). Because the number of population health items is small relative to the CLEO items, MCCQE Part I scores are reported as separate scores, as CLEO and PHELO (Population Health + CLEO). In effect, CLEO is a subscore of PHELO; both emphasise non-cognitive domains to a greater extent than the 5 other MCCQE Part I scores (A-P Boulais, Manager MCCQE Part I, Medical Council of Canada; personal communication 2006).

This study extends the findings of Eva *et al.* (2004)^{6,7} by comparing performance on the various admissions measures used by McMaster University's medical school with in-programme performance during clerkship and on the MCCQE Part I post-graduation. The following questions were asked:

- 1 Does the MMI predict clinical clerkship performance?
- 2 Does the MMI predict national licensing examination performance?
- 3 How does the predictive validity of the MMI compare with that of more traditional admission measures of professional qualities?

4 How does the predictive validity of the MMI compare with that of the uGPA?

METHODS

Participants

A total of 117 candidates on the undergraduate MD programme (30% of all interviewees) participated in the MMI in spring 2002. Of these, 45 were subsequently enrolled, making up 33% of the class of 2005; 43 students sat the MCCQE Part I in April 2005 and 2 in April 2006 (because they did not complete the 3-year undergraduate programme until then). All but 1 agreed to release their MCCQE Part I results. As the scores on the MCCQE I are transformed to have a constant mean and standard deviation each year, all participants were lumped into a single analysis. Excluding the 2 students who did not sit in 2005 from the analyses did not alter the conclusions. Consent to participate in the study was obtained from all subjects at the time of MMI administration.

Admissions tools

A total of 3027 candidates submitted an application package to the Ontario Medical School Application Service in October 2001 in an attempt to gain entry into McMaster University's class of 2005. Applicants were screened for invitation to interview using an equal weighting of GPA and scores assigned to an autobiographical submission (ABS). The top 384 candidates were invited to interview in March/April 2002.

In addition, each of the 384 interviewed candidates was invited to participate in an MMI. The invitation made it clear that volunteering, or declining to volunteer, for the MMI would in no way influence chances of admission. The first 120 volunteers were scheduled for 3 sequential MMI sessions, scheduled daily for all weekend interview dates, with all MMI sessions commencing only after completion of the standard interview processes. Three of the 120 volunteers withdrew due to illness on the dates of interview. The remaining 117 candidates completed a 10-station MMI. Each station lasted 8 minutes and was rated by a single examiner. Examiners were asked: 'Please rate the applicant's overall performance on this station', using a 7-point scale where 1 = unsatisfactory, 3 = borderline, 5 = satisfactory and 7 = outstanding.⁶

The same rating scale was used for the standard admission processes, which, like the MMI, were all

designed to assess candidates' non-cognitive (professional) tendencies (e.g. communication skills and moral reasoning). These standard processes included an ABS, a traditional interview, and a simulated tutorial. The ABS comprised a series of 15 short-answer format questions, to be completed remotely by all applicants in a non-invigilated setting. Each submitted ABS was rated independently by 3 trained raters, representing members of faculty, the medical student body and the community. Inter-rater reliability has varied from year to year for the ABS, with 1 set of results published as 0.45.⁵ The traditional interview and simulated tutorial were conducted on the interview date. The traditional interview involved the use of standard questions over a 45-minute period, asked by 3 interviewers representing faculty, students and community. Although the interview was conducted by all 3 interviewers at the same time, the interviewers scored the applicant independently, with a published inter-rater reliability of 0.66.⁵ Three independent raters representing faculty, students and community also observed groups of 6 applicants discussing 2 ethical issues as part of the simulated tutorial, with each of the 3 independent raters providing 2 separate ratings for each of the 2 challenges. Inter-rater reliability has been published as 0.61,⁵ but an earlier study found test-retest reliability of the simulated tutorial to vary between 0.09 and 0.18, depending on the phrasing of the rating scale question.¹¹

In-programme evaluation tools

All students in the undergraduate MD programme write a personal progress inventory (PPI) 8 times during their 3 years in the programme. The PPI is a 180-item MCQ examination that is intended to provide a broad-based measure of medical knowledge¹² and has a test-retest reliability in the 0.6–0.8 range.¹³

Objective structured clinical examinations are also used by the MD programme to assess clinical skills, including non-cognitive competencies of communication skills and problem-solving ability. Students complete a 10-station OSCE during each year of the programme, each of which has an interstation reliability typically in the 0.5 range (unpublished data).

For a 12-month period corresponding to the calendar year 2004, students in the class of 2005 completed their studies in 6 core clerkship rotations (family medicine, internal medicine, obstetrics and gynaecology, paediatrics, psychiatry, and surgery). Encounter cards are used in each rotation to collect information regarding students' clinical skills and professional behaviour; the

reliability of this tool is > 0.8 .¹⁴ In addition, clerkship directors complete an end-of-rotation assessment for each student. The end-of-rotation assessment is a qualitative summary of performance with quantitative ratings (using 10-point scales) on each of the following dimensions: fund of knowledge; knowledge integration; history taking; clinical examination; clinical management; learning skills; communication skills; respect; responsibility, and self-awareness. For the purpose of this study, averaging across each of the 10 ratings and each of 4 clerkships created a clerkship 'score'. (Ratings were unavailable for the first 2 clerkship rotations completed by these students because the scoring system changed from an entirely qualitative system to that described above in May of 2004.)

In addition, students can receive unsatisfactory or provisional satisfactory ratings that result in remediation for professional behaviour issues from any clerkship rotation director or pre-clerkship tutor.

MCCQE Part I

The MCCQE Part I was described in the introduction to this paper. The Medical Council of Canada reports separate scores for medicine, surgery, obstetrics and gynaecology, psychiatry, paediatrics, CLEO and PHELO, as well as the composite scores for the MCQ component taken in the morning and the Key Features clinical decision-making component in the afternoon. For the purpose of our analyses, we have kept each subscore separate in order to test whether or not different admissions measures were differentially predictive of different aspects of licensing examination performance.

Analysis

We examined the correlation between admissions measures, in-course measures, and the MCCQE Part I

using Pearson's correlation coefficients. Regression analyses were then performed to determine which admissions tools were statistically predictive of each outcome when scores on the other admissions tools were taken into account (i.e. to determine the independent predictability of each admissions tool).

RESULTS

As indicated by Eva *et al.*, all the evaluation data collected for the sample of students included in this study indicate that this cohort was representative of the MD programme's class of 2005.⁷ Table 1 illustrates the correlation between each admissions tool and various intramural assessment exercises. Inclusion of all 5 admissions tools in a series of regression analyses confirmed that:

- 1 only the MMI was statistically predictive of OSCE performance (standardised $\beta = 0.4$, $P < 0.05$);
- 2 GPA and the simulated tutorial were both predictive of PPI performance (standardised $\beta = 0.3$, $P < 0.06$), although the relation between the simulated tutorial and PPI was negative, indicating that those who performed worse on the simulated tutorial performed better on the PPI, and
- 3 only the MMI was statistically predictive of clerkship performance, measured with both the average ratings assigned by clerkship directors (standardised $\beta = 0.7$, $P < 0.001$) and encounter card ratings provided by clinical preceptors (standardised $\beta = 0.5$, $P < 0.01$).

Table 2 illustrates the correlation between each admissions tool and performance on the subscales of the MCCQE Part I. All 5 admissions tools were submitted to a series of regression analyses to determine which were independently predictive of Part I results. These analyses revealed that:

Table 1 Correlations between admissions tools and in-programme assessment exercises

	Objective structured clinical examination	Personal progress inventory	Clerkship director ratings	Clerkship encounter cards
Autobiographical submission	0.02	- 0.14	- 0.03	0.26
Grade point average	0.08	0.33†	- 0.06	- 0.24
Multiple Mini-Interview	0.28*	0.04	0.57‡	0.51‡
Personal interview	0.15	- 0.07	0.06	0.12
Simulated tutorial	- 0.17	- 0.28*	0.04	0.11

* $P < 0.10$; † $P < 0.05$; ‡ $P < 0.01$

Table 2 Correlations between admissions tools and performance on Part I of the Medical Council of Canada's Qualifying Examination (MCCQE)

	CLEO	Medicine	Obs/Gyn	Paediatrics	PHELO	Psychiatry	Surgery	MCQ	CDM	Total score
Autobiographical submission	-0.08	-0.11	-0.02	-0.05	0.00	0.02	0.06	-0.05	-0.11	0.00
Grade point average	0.04	0.30†	0.15	0.35†	0.05	0.24*	-0.01	0.30†	0.25*	0.26*
Multiple Mini-Interview	0.39‡	-0.04	0.02	0.03	0.37‡	0.08	0.06	0.10	0.18	0.17
Personal interview	-0.11	0.23	0.20	-0.05	-0.27*	-0.05	-0.02	0.07	0.08	-0.04
Simulated tutorial	0.05	-0.2*	-0.18	-0.20	0.16	-0.03	-0.11	-0.19	-0.25*	-0.28*

CLEO = Considerations of the Legal, Ethical and Organisational Aspects of Medicine

Obs/Gyn = obstetrics and gynaecology

PHELO = Population Health and Ethical, Legal and Organisational Aspects of Medicine

MCQ = multiple-choice question component of MCCQE Part I

CDM = clinical decision-making component of MCCQE Part I

* $P < 0.10$; † $P < 0.05$; ‡ $P < 0.01$

- 1 only the MMI was statistically predictive of CLEO or PHELO performance (standardised $\beta > 0.4$, $P < 0.01$);
- 2 only uGPA was statistically predictive of MCQ performance (standardised $\beta = 0.38$, $P < 0.05$);
- 3 only the MMI was predictive of CDM performance (standardised $\beta = 0.35$, $P < 0.05$), and
- 4 the MMI and uGPA were equally predictive of overall test performance (standardised $\beta > 0.3$, $P < 0.06$).

Undergraduate uGPA was correlated with medicine, paediatrics, and psychiatry subscores, but none of our admissions tools were predictive of obstetrics and gynaecology or surgery subscores.

One of the 45 students in this sample received an unsatisfactory mark for professional behaviour during a pre-clerkship unit, as reported by Eva *et al.*⁷ Since then, a second student received a similar unsatisfactory mark during a clerkship rotation. As these are the extreme cases of most concern for medical admissions personnel, it is worth considering how each student was rank ordered by the various admissions tools, although conclusive judgements should not be drawn from ratings of only 2 individuals. Of the 45 students included in the sample, the MMI ranked these 2 students lowest (in the 15th and 11th percentiles, respectively) relative to the other 4 admissions tools. The percentiles in which these students fell according to the simulated tutorial were the 76th and 93rd; according to GPA these students were in the 76th and 80th percentiles; according to the personal interview they fell in the 48th and 27th percentiles; and, finally, the autobiographical submission ranked these students in the 17th and 64th percentiles. In other words, the MMI was most consistently likely to have excluded these 2 students when admissions decisions were made.

DISCUSSION

The predictive validity of cognitive measures is well established. Undergraduate academic achievement as measured by uGPA and aptitude test scores like MCAT predict for performance at pre-clerkship and clerkship levels of medical school, as well as on national licensing examinations.² A literature review concluded that most studies identify a positive relationship between US Medical Licensing Examination (USMLE) scores during medical school and subsequent ratings of residents' clinical performances and/or scores on specialty board certification examinations.¹⁵ Certification examination results predict post-training medical practice as measured by several end-points: appropriate use of radiological investigations;^{16,17} acceptable patterns of drug prescription;^{16,17} mortality rates following cardiac events,¹⁸ and peer review.¹⁹

Demonstrating predictive validity in the non-cognitive, professional quality domain has been more difficult. Relatively few measures of non-cognitive professional qualities collected during medical school or residency meet the necessary criterion of reliability.²⁰ Over the last decade, however, access to richer descriptions of professional qualities, in keeping with the recommendations of Ginsburg *et al.*,²¹ has become available. Increasingly, positive correlations are being drawn across the time periods of pre-admission to medical school, during medical school, at national licensure examinations, during residency training and during post-training practice. Empathy scores in medical school correlate with ratings of empathic behaviour in residency training 3 years later.²² In an elegant study, records of investigations of disciplinary action against practising doctors²³ were found to be predictable on the basis of narrative assessments by raters provided during

medical school. Based upon the results of this study, the MMI can be added to that chain of evidence, as it predicts clerkship performance scores and success on professional domains of national licensing examinations. Although there are fewer correlations than there are relating to cognitive measures, the emerging picture is nevertheless increasingly compelling.

Some limitations of this study need to be acknowledged. The sample size of the study was limited. However, one would expect that this would make the study more susceptible to type II, rather than type I, error. The predictive validity correlations of the MMI with in-school and licensure scores were moderately strong at 0.28–0.57, comparable with published results for cognitive measures of GPA (0.26–0.40) and MCAT (0.32–0.61).² Had the MMI results failed to demonstrate statistical significance, the limited sample size could have been cited as causative; demonstrating statistical significance despite the limited sample size is all the more impressive. Further, in an ideal world, a medical school would conduct a study in which all interviewed applicants went through traditional procedures, as well as MMIs, providing improved power of a full-class cohort, as well as comparisons with historical admissions evaluations. The resource implications of this approach are sufficiently daunting as to raise the possibility that such a study might never be completed.

As a further potential limitation, the selection process did not use the MMI; however, as indicated by Eva *et al.*, the coefficient of variation did suggest that the MMI and other admissions measures suffered equally from restriction of range.⁷ The MMI is now a central component of the admissions process at McMaster and, after 2 years, we have yet to witness a downwards drift of MMI reliability, thereby further lessening concern about this limitation, but we will continue to monitor the MMI's predictive validity.

Further maturation of this study cohort's data will allow assessment of the correlation between MMI scores and MCCQE Part II, the OSCE-based component of the Canadian national licensing examination that is typically completed 16 months into residency.²⁴ Further, recent implementation of the MMI at other medical schools that, unlike McMaster, use the MCAT, will allow comparisons heretofore limited to uGPA.

CONCLUSIONS

In follow-up of the first large MMI study, the data have matured only recently to the point of enabling

assessment of the correlations with MCCQE Part I, and, in particular, the CLEO and PHELO subsections. Additionally, the first study cohort of MMI-experienced medical students has completed all intramural evaluations at the Michael G DeGroot School of Medicine at McMaster University, including professional quality measures of OSCE, and clerkship evaluations. These data therefore provide answers to the research questions asked here.

- 1 Does the MMI predict for clinical clerkship performance? Yes. Statistically significant and moderately large positive correlations were found.
- 2 Does the MMI predict for national licensing examination performance? Yes. Statistically significant positive correlations were found, but only for those subsections assessing professional quality domains.
- 3 How does the predictive validity of the MMI compare with that of more traditional admission measures of professional qualities? Favourably. More traditional admissions measures of professional qualities, including panel-style personal interviews, simulated tutorials, and autobiographical submissions were found *not* to provide any predictive validity for subsequent scores during clinical clerkship or on Part I of the Canadian national licensing examination.
- 4 How does the predictive validity of the MMI compare with that of uGPA? Comparably. The correlation of uGPA with core clerkship subscores on the national licensing examination and the correlation of MMI scores with professional quality domain scores on the same examination were of similar magnitude. Undergraduate GPA was better for predicting progress testing in medical school, but the MMI was better for predicting both OSCE and clerkship ratings, suggesting that uGPA and MMI function best in complementary roles to each other.

Contributors: all authors contributed to the conception and design of the study. KWE collected the data and ran the analyses. HIR drafted the manuscript. KWE, JR and GRN made critical revisions. All authors approved the final version.

Acknowledgements: none.

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 provides no official endorsement.

Conflicts of interest: none.

Ethical approval: this study was approved by the Hamilton Health Sciences Research Ethics Board.

REFERENCES

- 1 Albanese MA, Farrell P, Dottl S. Statistical criteria for setting thresholds in medical school admissions. *Adv Health Sci Educ* 2005;**10**:89–103.
- 2 Julian ER. Validity of the Medical College Admission Test for predicting medical school performance. *Acad Med* 2005;**80**:910–7.
- 3 Albanese MA, Snow MH, Skochelak SE, Huggett KN, Farrell PM. Assessing personal qualities in medical school admissions. *Acad Med* 2003;**78**:313–21.
- 4 Salvatori P. Reliability and validity of admissions tools used to select students for the health professions. *Adv Health Sci Educ* 2001;**6**:159–75.
- 5 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.
- 6 Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Med Educ* 2004;**38**:314–26.
- 7 Eva KW, Reiter HI, Rosenfeld J, Norman GR. The ability of the multiple mini-interview to predict pre-clerkship performance in medical school. *Acad Med* 2004;**79** (Suppl):40–2.
- 8 Papadakis MA, Loeser H, Healy K. Early detection and evaluation of professionalism deficiencies in medical students: one school's approach. *Acad Med* 2001;**76**:1110–6.
- 9 Page G, Bordage G. The Medical Council of Canada's key features project: a more valid written examination of clinical decision-making skills. *Acad Med* 1995;**70**:104–10.
- 10 Page G, Bordage G, Allen T. Developing key-feature problems and examinations to assess clinical decision-making skills. *Acad Med* 1995;**70**:194–201.
- 11 Reiter HI, Rosenfeld J, Giordano L. Selection of medical students at McMaster University: a quarter century later. *McMaster University Med J* 2004;**2**:41–5.
- 12 Blake JM, Norman GR, Keane DR, Mueller CB, Cunningham J, Didyk N. Introducing progress testing in McMaster University's problem-based medical curriculum: psychometric properties and effect on learning. *Acad Med* 1996;**71**:1002–7.
- 13 Eva KW, Cunningham JPW, Reiter HI, Keane DR, Norman GR. How can I know what I don't know? Poor self-assessment in a well defined domain. *Adv Health Sci Educ* 2004;**9**:211–24.
- 14 Hatala R, Norman GR. In-training evaluation during an internal medicine clerkship. *Acad Med* 1999;**74** (Suppl):118–20.
- 15 Berner ES, Brooks CM, Erdmann JB. Use of the USMLE to select residents. *Acad Med* 1993;**68**:753–9.
- 16 Tamblyn R, Abrahamowicz M, Dauphinee WD, Hanley JA, Norcini J, Girard N, Grand'Maison P, Brailovsky C. Association between licensure examination scores and practice in primary care. *JAMA* 2002;**288**:3019–26.
- 17 Tamblyn R, Abrahamowicz M, Brailovsky C, Grand'Maison P, Lescop J, Norcini J, Girard N, Haggerty J. Association between licensing examination scores and resource use and quality of care in primary care practice. *JAMA* 1998;**280**:989–96.
- 18 Norcini JJ, Lipner RS, Kimball HR. Certifying examination performance and patient outcomes following acute myocardial infarction. *Med Educ* 2002;**36**:853–9.
- 19 Ramsey PG, Carline JD, Inui TS, Larson EB, LoGerfo JP, Wenrich MD. Predictive validity of certification by the American Board of Internal Medicine. *Ann Int Med* 1989;**110**:719–26.
- 20 Veloski JJ, Fields SK, Boex JR, Blank LL. Measuring professionalism: a review of studies with instruments reported in the literature between 1982 and 2002. *Acad Med* 2005;**20**:151–7.
- 21 Ginsburg S, Regehr G, Hatala R, McNaughton N, Frohna A, Hodges B, Lingard L, Stern D. Context, conflict, and resolution: a new conceptual framework for evaluating professionalism. *Acad Med* 2000;**75** (Suppl):6–11.
- 22 Hojat M, Mangione S, Nasca TJ, Gonnella JS, Magee M. Empathy scores in medical school and ratings of empathic behavior in residency training 3 years later. *J Soc Psychol* 2005;**145**:663–72.
- 23 Papadakis MA, Teherani A, Banach MA, Knettler TR, Rattner SL, Stern DT, Veloski JJ, Hodgson CS. Disciplinary action by medical boards and prior behavior in medical school. *New Engl J Med* 2005;**353**:2673–82.
- 24 Smee SM, Dauphinee WD, Blackmore DE, Rothman AI, Reznick RK, Des Marchais J. A sequenced OSCE for licensure: administrative issues, results and myths. *Adv Health Sci Educ* 2003;**8**:223–36.

Received 28 July 2006; accepted for publication 31 October 2006