Selection into medicine using interviews and other measures: Much remains to be learned

Colleen Ma
The University of Sydney
Peter Harris, Andrew Cole, Phil Jones and Boaz Shulruf
University of New South Wales

The objectives of this study were to identify the effectiveness of the panel admission interview as a selection tool for the medical program and identify improvements in the selection tools battery. Data from 1024 students, representing four cohorts of students were used in this study. Exploratory factor analysis using principal component analysis was used to identify underlying factors within the admission tools. A series of hierarchical linear regressions was employed to identify the predictability of performance in the medical program by the admission tools. Although the admission tools yielded low correlations with one another (r<.30), correlations between interview sub-scores were high (.435<r<.640). All interview sub-scores loaded on to a single factor explaining over 60% of the variance. The admission tools and the interview overall scores explained less than 13.5% and 3.8% (respectively) of the variance in the key outcome measures. We concluded that each admission tool measured different attributes, and suggest that admission interview procedures and the interview questions should be assessed independently.

Introduction

It is common experience that the number of medical school applicants greatly exceeds the number of available places in each program (Barzansky & Etzel, 2003). Therefore, the admission process plays an important role in helping to identify those students with the desired skills and attributes to be successful in the medical program. The selection of students involves the use of a combination of admission tools in order to determine the most suitable candidates. Some institutions also aim to increase the diversity of their medical students measured by different social determinants (Lakhan, 2003; Puddey & Mercer, 2013) which have been found to impact on student self-concept, learning and later, on their career choices (Poole, Bourke & Shulruf, 2010; Yeung, Li, Wilson & Craven, 2013). It is noteworthy that medical school selection tools are rarely designed to correlate with particular assessments within the medical program (Eva, Reiter, Rosenfeld, Trinh, Wood & Norman, 2012). Admission tools that are widely used around the world consider previous grades (GPA), aptitude or achievement tests, interviews, reference letters and personal essays (Adam, Dowell & Greatrix, 2011; AAMC, 2017; Eva, Reiter, Rosenfeld, Trinh, Wood & Norman, 2012; Mercer & Puddey, 2011; Prideaux et al., 2011; Shulruf, Poole, Wang, Rudland & Wilkinson, 2012; Wilkinson, Zhang & Parker, 2011; Wright & Bradley, 2010). Despite the plethora of research in this area, predicting performance in the medical program, including timely completion, remains a major challenge (Shulruf et al., 2012).

Previous academic performance, measured by GPA, was found to be the strongest predictor of subsequent academic success and future career placement, yet the predictability of academic success is stronger for the early years and drops towards the end of the program (Cohen-Schotanus et al., 2006; Collins & White, 1993; Dowell et al., 2011; McManus et al., 2005; Shulruf, Poole, Wang, Rudland & Wilkinson, 2012; Silver & Hodgson, 1997; Wilkinson et al., 2008). Aptitude and achievement tests such as the Undergraduate Medicine and Health Sciences Admissions Test (UMAT), Medical College Admissions Test (MCAT) and the UK Clinical Aptitude Test (UKCAT) are also becoming prevalent in the medical program admissions process. Of the three mentioned, both the UMAT and UKCAT are considered to be purely aptitude tests, while some sections of the MCAT (which are used for graduate entry programs) are deemed to measure academic knowledge previously acquired (Prideaux et al., 2011).

UMAT scores have been shown to be a weak predictor of academic performance in medical school (Mercer, Abbott, & Puddey, 2012; Shulruf et al., 2012; Wilkinson et al., 2011). Likewise, the UKCAT has also been shown to be a poor predictor of study success after admission (Lynch, MacKenzie, Dowell, Cleland & Prescott, 2009). Aptitude and achievement tests were found to interact with ethnicity and socioeconomic factors, which suggest that additional admission tools are required (Davis et al., 2013; Puddey & Mercer, 2013).

Some medical schools utilise interviews as an important part of the selection process, aiming to assess a broader range of a candidate's attributes, such as interpersonal skills, motivation and personality, that are not as readily assessed through GPA and admission tests scores (Albanese, Snow, Skochelak, Huggett & Farrell, 2003). For example, a recent study from Australia reported that removing interviews from the selection process was associated with gender bias as it increased the proportion of males being admitted to the medical program (Wilkinson, Casey & Eley, 2014). Nonetheless, the research on admission interviews suggests that interviews only poorly predict future performance, academic or otherwise (Prideaux et al., 2011; Shulruf et al., 2012; Wilkinson et al., 2008).

Moreover, Kulatunga Moruzi and Norman (2002) suggested that despite achieving an acceptable (0.66) inter-rater reliability for the overall rating of the admission interviews, there was no significant relation between interview scores and performance in clinical tasks. The only exception to this is the Multiple Mini Interview (MMI) which utilises a format similar to that of the OSCEs and shows a significant relationship with success on OSCE performance during the clinical years, probably due to the similarity between those two assessments (Eva et al., 2012; Pau et al., 2013; Reiter, Eva, Rosenfeld & Norman, 2007).

In Australia, ten schools offer undergraduate entry medical programs, with the majority of students entering directly after secondary school graduation. Applicants are required to undergo an assessment process comprising a structured interview, an aptitude test (UMAT), a high secondary school GPA and a rural score if applicable (Monash University, 2014; University of New South Wales, 2012; University of Western Australia, 2014).

UNSW medical school makes use of an integrated selection process which takes into account the items mentioned previously. Part of this process includes the interview, from which a recent study showed a strong relationship between its communication dimension and clinical competency in the medical program (Simpson et al., 2014). The interview process in our medical school involves a structured 40 minute interview by two interviewers who each score the interview independently at the its conclusion and are then required to reach a consensus score. On the rare occasions this consensus is not reached a later second interview is needed. Interviewers are required to undergo training every two years to recalibrate. This interview process was designed and implemented two years before the curriculum change and independently of the associated assessment system. In the interview, proficiency is considered across six predefined domains: communication skills, motivation, empathy towards others, self-awareness, responding to diversity and ability to cope with uncertainty (Simpson et al., 2014).

The aim of this study was to identify the ability of the admission tools to predict medical student performance in later core assessment tasks within the medical program.

Methods

Data

The data for this study included predictor variables such as ATAR (Australian Tertiary Admission Rank), UMAT (a selection tool consisting of three domains: "logical reasoning and problem solving", "understanding people" and "non-verbal reasoning") and interview scores for all accepted candidates that went on to complete the medical program (University of New South Wales, 2012). Performance data from a number of non-clinical and clinical examination results from various stages of the six year medical program were included in the study. For those students who had to repeat an assessment, the original result was used for the purposes of this study.

Analysis

Analysis of the data was carried out using *SPSS* (version 22). 95% confidence intervals of the means of ATAR, UMAT and the admissions interview were calculated for each of the cohorts. Correlation between the three admissions tools (ATAR, UMAT and interview) were analysed by calculating Pearson correlation coefficients.

Exploratory factor analysis (EFA) (employing principal component analysis) was then conducted to identify any discrete factors within the six expected domains of the interview questions. Multiple linear regression analysis was used to determine the predictability of students' examination scores throughout the three phases of the medical program by the interview scores.

Ethics

The study was approved by the Medical and Community Human Research Ethics Advisory Panel (ref 2013-7-51).

Results

Descriptive statistics

Data from 1047 students were available but records of 20 students were not complete, thus only 1027 (98%) were used in the analyses (2004 N=251 females=57.4% ; 2005 N=197 females=56.1%; 2006 N=271 females=55.6%; 2007 N=308 Females=56.9%). Analysis of the four different cohorts showed no significant differences across the cohorts for ATAR, the three UMAT sections and for the six sections of the interview (Table 1).

Table 1: 95% Confidence Intervals of the mean score of selection tools by cohorts

Tools	Cohort			
	2004	2005	2006	2007
ATAR	98.73-99.14	98.12-98.72	98.75-99.12	98.63-99.68
UMAT Sec 1	61.12-63.31	60.71-63.08	58.37-60.28	59.27-61.43
UMAT Sec 2	66.64-69.94	57.66-60.40	56.68-58.72	56.10-57.98
UMAT Sec 3	62.64-64.88	61.11-64.29	61.38-63.89	61.97-64.28
Interview Sec 1	5.84-6.04	5.56-5.81	5.65.-5.85	5.53-5.73
Interview Sec 2	5.84-6.06	5.71-5.98	5.69-5.91	5.54-5.74
Interview Sec 3	5.72-5.92	5.57-5.81	5.63-5.84	5.45-5.64
Interview Sec 4	5.59-5.81	5.50-5.77	5.51-5.72	5.35-5.54
Interview Sec 5	5.75-5.98	5.76-6.00	5.69-5.91	5.59-5.79
Interview Sec 6	5.65-5.88	5.57-5.83	5.48-5.72	5.39-5.59

Exploratory factor analysis (principal component analysis) for all six domains of the interview (each cohort separately) identified only a single factor, which explained over 60% of the variance (variance explained by cohort: 2004 – 61.3%; 2005 – 64.7%; 2006 – 63.4%; 2007 – 61.0%).

Table 2: Correlations between admission tools

	ATAR	UMAT1	UMAT2	UMAT3	INT1	INT2	INT3	INT4	INT5
UMAT1	.168**								
UMAT2	.018	.182**							
UMAT3	.284**	.291**	-.029						
INT1	-.154**	-.113**	.083**	-.146**					
INT2	-.061*	-.060	.115**	-.143**	.559**				
INT3	-.109**	-.093**	.075*	-.135**	.630**	.564**			
INT4	-.128**	-.091**	.101**	-.172**	.549**	.640**	.606**		
INT5	-.100**	-.105**	.038	-.123**	.473**	.481**	.480**	.435**	
INT6	-.117**	-.114**	.061*	-.134**	.577**	.593**	.581**	.594**	.509**

*p<.05; **p<.01

In order to identify associations across the admission tools used, Pearson correlations were calculated between each of them (ATAR; UMAT 1-3; and scores of interview domains 1-6). The results demonstrate that ATAR was not highly correlated with any of the other measures. The highest correlation was with UMAT3 ($r=.284$ $p<.01$); UMAT 1-3 scores had low correlations with each other, while UMAT2 scores did not correlate with the other UMAT scores; UMAT 1-3 did not have any meaningful correlations with any of the interview domains' scores, which did have relatively high correlations with each other (.435<r<.640, p<.01) (Table 2).

Table 3: Variance explained by selection tools by measured outcomes (summary results from linear regressions)

Year	Measured outcomes**	Regression model*			Var 1	Var 2
		Model 1	Model 2	Model 3		
2	Clinical communication skills	3.1%	3.6%	7.5%	3.9%	4.4%
	End of phase exam	18.6%	31.4%	32.1%	0.7%	13.5%
	Portfolio	2.6%	8.9%	9.4%	0.5%	6.8%
4	Integrated clinical assessment	1.1%	3.9%	6.2%	2.3%	5.1%
	Portfolio	4.5%	8.9%	10.8%	1.9%	6.3%
6	Biomedical sciences viva	2.6%	8.3%	9.7%	1.4%	7.1%
	Integrated clinical exam	4.1%	5.1%	6.7%	1.6%	2.6%
	Portfolio	1.7%	3.0%	3.9%	0.9%	2.2%

* Model 1: Predictors: demographic variables
Model 2: Predictors: demographics and academic achievement
Model 3: Predictors: demographics and academic achievement and interview
Var 1: Total variance explained by interview scores
Var 2: Total variance explained by all admission tools

** Measured outcomes:

- Clinical communication skills: ability to effectively communicate in clinical setting
- End of phase exam: results of a written medical knowledge test summarising knowledge of the two years of each phase
- Integrated clinical assessment: performance in examination of which students demonstrate their clinical knowledge and skills.
- Biomedical sciences viva: oral examination focusing on biomedical sciences knowledge
- Portfolio: each student prepares a portfolio of their clinical and biomedical learning activities and performance. The portfolio is assessed by oral examination at the end of each phase (year 2, 4, 6)

To identify the predictability of key outcomes in the medical program by the three admission tools, a series of hierarchical multiple linear regressions models was employed. The models used three blocks: block 1 demographic variables (gender, cohort); block 2 ATAR and UMAT scores; and block 3 interview scores. Table 3 presents the outcomes' variance explained for each of the models. Overall the admission tools did not predict the key outcomes (Table 3). Altogether, the admission tools explained 13.5% of the variance in end of Year 2 written examination scores, followed by 7.1% of the variance in Year 6 (final) Biomedical Sciences Viva scores, 6.8% and 6.3% of the variance in Year 2 and Year 4 (respectively) Portfolio scores. The interview scores explained only 3.9% of the variance in Year 2 Clinical Communication Skills, 2.3% in the variance of Year 4 Integrated Clinical Assessment and 1.6% of the variance in Year 6 Integrated Clinical Examination (Table 3).

Discussion

The aim of this study was to identify the effectiveness of the admissions interview as a tool for predicting candidates' assessment performance in the medical program. Eight key outcomes (Table 3) were measured in this study that enabled us to identify the impact of the interview scores on achievement throughout the medical program. The results clearly indicate that the admission interview scores explained very little variance in any measured key outcome (Table 3). This finding echoes previous studies conducted in other institutes in Australia and elsewhere (Lumb, Homer & Miller, 2010; Mercer & Puddey, 2011; Poole, Shulruf, Rudland & Wilkinson, 2012; Prideaux et al., 2011; Salvatori, 2001; Shulruf et al., 2012). The question is therefore why selection interviews, which require a lot of resources, explain so little variance in student performance throughout the program? For example, in our institute each interview takes about 45 minutes and employs 2 interviewers and about half of the interviewees are admitted. There are a number of plausible explanations for this phenomenon.

The first explanation is that the selection process had been very successful. Generally the dropout from medical programs is low (O'Neill, Wallstedt, Eika, & Hartvigsen, 2011). For example reports form the UK suggest 3.8 to 4.2% dropout rate (Arulampalam, Naylor & Smith, 2004, 2007) and a recent study from our institute reported less than 2% of year 1 discontinuation of students who studied in the program, which is about a half of the rate observed before the current admission process was put in place (Simpson et al., 2014). Thus it is possible that very few students selected were not suitable for the program, which means that the selection tools worked well, as intended in distinguishing between the suitable and non-suitable applicants rather than predicting achievement within the program. This could explain the low predictive power of any of the admission tools including the interviews. It is noted that similar correlations between admission interview scores with clinical and communication skills assessed later in the program were found in studies undertaken previously on similar but not identical populations (Mercer, 2007; Mercer et al., 2012; Puddey, Mercer, Carr, & Louden, 2011; Simpson et al., 2014).

An alternative explanation is that the interviews were not efficient enough. Although aimed to measure six discrete domains, our analysis suggests that the interview scores all measured the same trait. The correlations between the domain scores were high (Table 3) All the scores loaded on to a single factor which explains about 60% of the variance. Given that the interview schedule was designed by a professional team to measure different traits, it is possible that the first impression the interviewee made on the interviewers was the strongest, or any other particular strong impression that overshadowed responses to most of the questions asked throughout the interview (McLaughlin, 2014; Wood, 2014).

It is also suggested that more general issues related to reliability, and predictive validity of admission interviews, which have been widely reported, may have impacted on the effectiveness of the interviews undertaken in our institution as well (Edwards, Johnson, & Molidor, 1990; Lumb et al., 2010; Poole, Shulruf, Harley, et al., 2012; Salvatori, 2001). A possible avenue for improvement might be employing a mini multiple interview (MMI) technique which has been reported to yield better predictive validity, particularly predicting performance in clinical skills assessments (Eva et al., 2012; Pau et al., 2013). The MMI is a series of sequential short interviews each of which focuses on a particular set of skills and each is conducted by a single interviewer.

Therefore, possible ways to improve the predictive validity of the admission interviews in our institution might be by splitting the panel interview to six MMI stations, each measuring one of the domains as currently intended to be assessed. Using this practice may provide some more insight into the admission interview process. A comparison of the predictive validity of the suggested MMI with the currently used panel interview may identify the impact of 'first impression' on the interview results (McLaughlin, 2014; Wood, 2014). This is a low risk change as it only requires operational change without changing the content of the interview questions. Given the low risk, this practice could be applicable for any medical program that currently utilises a similar admission interview process and might enable those programs to make better informed decisions about which way to go in the future. This approach will not address other "softer" outcomes nor issues of career selection.

The other important finding of this study is the low correlations that were found between the different selection tools (Table 3). Although similar findings had been reported previously, the issue of such low correlations has been scarcely discussed in detail (Basco, Lancaster, Gilbert, Carey & Blue, 2008; Carr, 2009; Kulatunga Moruzi & Norman, 2002). If selection tools did not correlate but were found to provide reliable (not implying validity here) measures, then each tool may be deemed to measure a discrete trait, or a set of attributes, different from the others.

Given that medical professional practice is comprised of different sets of skills and qualities, it is suggested that admission tools' validity be measured by comparing each admission tool separately against its corresponding attribute as manifested within the medical school assessment schedules. Applying such a student selection policy may bring some new opportunities for the medical workforce. Different medical specialties require different strengths (Harrold, Field & Gurwitz, 1999; Smetana et al., 2007). Our literature search did not identify any medical program that applied a differential admission policy based on forecast medical workforce needs. A recent study from New Zealand (Poole & Shulruf, 2013) identified that medical school applicants who had strong interest in general practice (GP) scored between 3-5 points lower on UMAT tests ($p<.02$) than those who did not have interest in GP. Interestingly, admission GPA and interview scores did not differ across those groups. Such findings demonstrate that applying an admission policy of 'one size suits all' may not be the most efficient in fulfilling society's needs. The Consensus statement and recommendations from the Ottawa 2010 Conference (Prideaux et al., 2011) alluded to this by recommending a focus on multi-method programmatic approaches which are fit for purpose while considering medical schools' social accountability in relation to social inclusion and workforce issues.

It is acknowledged that this study has some limitations. The major limitation was the availability of data, particularly the lack of information on the interviewers. Those data were not available and therefore it was impossible to measure inter-rater reliability. Another limitation is that the study included only students who have completed the program. No data from those who were not admitted to the program or dropped out were analysed. This is a common limitation in similar studies undertaken within a single institute and no remedy could be offered unless the measured outcome includes discontinuation (Callahan, Hojat, Veloski, Erdmann & Gonnella, 2010; Shulruf et al., 2012) or includes multi-institutional data where applicants who had not been admitted to one institute could be admitted to others (Kaur, Roberton & Glasgow, 2013).

Conclusions

Selecting the best candidates for a medical program has been a major challenge for many years. The current study might have not resolved many of the issues, yet it highlights a few avenues for further advancement in the field. In particular this study emphasises the need to measure the effectiveness of admission tools against a broad range of outcomes within and beyond the medical program.

References

AAMC (Association of American Medical Colleges) (2017). *2017 AMCAS instruction manual.* Washington DC: AAMC. https://aamc-orange.global.ssl.fastly.net/production/media/filer_public/c0/f8/c0f8833d-a302-46c7-b726-1b153dbac6de/2017_amcas_instruction_manual-_final.pdf

Adam, J., Dowell, J. & Greatrix, R. (2011). Use of UKCAT scores in student selection by UK medical schools, 2006-2010. *BMC Medical Education,* 11:98. http://dx.doi.org/10.1186/1472-6920-11-98

Albanese, M., Snow, M., Skochelak, S., Huggett, K. & Farrell, P. (2003). Assessing personal qualities in medical school admissions. *Academic Medicine,* 78(3), 313-321. http://journals.lww.com/academicmedicine/Fulltext/2003/03000/Assessing_Personal_Qualities_in_Medical_School.16.aspx

Arulampalam, W., Naylor, R. & Smith, J. (2004). Factors affecting the probability of first year medical student dropout in the UK: A logistic analysis for the intake cohorts of 1980-92. *Medical Education,* 38(5), 492-503. http://dx.doi.org/10.1046/j.1365-2929.2004.01815.x

Arulampalam, W., Naylor, R. & Smith, J. (2007). Dropping out of medical school in the UK: Explaining the changes over ten years. *Medical Education,* 41(4), 385-394. http://dx.doi.org/10.1111/j.1365-2929.2007.02710.x

Barzansky, B. & Etzel, S. I. (2003). Educational programs in US medical schools, 2002-2003. *JAMA (Journal of the American Medical Association),* 290(9), 1190-1196. http://dx.doi.org/10.1001/jama.290.9.1190

Basco, W., Lancaster, C., Gilbert, G., Carey, M. & Blue, A. (2008). Medical school application interview score has limited predictive validity for performance on a fourth year clinical practice examination. *Advances in Health Sciences Education,* 13(2), 151-162. http://dx.doi.org/10.1007/s10459-006-9031-5

Callahan, C., Hojat, M., Veloski, J., Erdmann, J. B. & Gonnella, J. S. (2010). The predictive validity of three versions of the MCAT in relation to performance in medical school, residency, and licensing examinations: A longitudinal study of 36 classes of Jefferson Medical College. *Academic Medicine,* 85(6), 980-987. http://www.ncbi.nlm.nih.gov/pubmed/20068426

Carr, S. E. (2009). Emotional intelligence in medical students: Does it correlate with selection measures? *Medical Education,* 43(11), 1069-1077. http://dx.doi.org/10.1111/j.1365-2923.2009.03496.x

Cohen-Schotanus, J., Muijtjens, A., Reinders, J., Agsteribbe, J., van Rossum, H. & van der Vleuten, C. (2006). The predictive validity of grade point average scores in a partial lottery medical school admission system. *Medical Education,* 40(10), 1012-1019. http://dx.doi.org/10.1111/j.1365-2929.2006.02561.x

Collins, J. P. & White, G. R. (1993). Selection of Auckland medical students over 25 years: A time for change? *Medical Education,* 27(4), 321-327. http://dx.doi.org/10.1111/j.1365-2923.1993.tb00276.x

Davis, D., Dorsey, K., Franks, R., Sackett, P., Searcy, C. & Zhao, X. (2013). Do racial and ethnic group differences in performance on the MCAT exam reflect test bias? *Academic Medicine,* 88(5), 593-602. http://www.ncbi.nlm.nih.gov/pubmed/23478636

Dowell, J., Lumsden, M. A., Powis, D., Munro, D., Bore, M., Makubate, B. & Kumwenda, B. (2011). Predictive validity of the personal qualities assessment for selection of medical students in Scotland. *Medical Teacher,* 33(9), e485-e488. http://dx.doi.org/10.3109/0142159X.2011.599448

Edwards, J., Johnson, E. & Molidor, J. (1990). The interview in the admission process. *Academic Medicine,* 65(3), 167-177. http://journals.lww.com/academicmedicine/Abstract/1990/03000/The_interview_in_the_admission_process_.8.aspx

Eva, K., Reiter, H. I., Rosenfeld, J., Trinh, K., Wood, T. & Norman, G. (2012). Association between a medical school admission process using the multiple mini-interview and national licensing examination scores. *JAMA (Journal of the American Medical Association),* 308(21), 2233-2240. http://dx.doi.org/10.1001/jama.2012.36914

Harrold, L., Field, T. & Gurwitz, J. (1999). Knowledge, patterns of care, and outcomes of care for generalists and specialists. *Journal of General Internal Medicine,* 14(8), 499-511. http://dx.doi.org/10.1046/j.1525-1497.1999.08168.x

Kaur, B., Roberton, D. M. & Glasgow, N. J. (2013). Evidence-based medical workforce planning and education: The MSOD project. *The Medical Journal of Australia,* 198(10), 518-519. http://dx.doi.org/10.5694/mja13.10243

Kulatunga Moruzi, C. & Norman, G. R. (2002). Validity of admissions measures in predicting performance outcomes: The contribution of cognitive and non-cognitive dimensions. *Teaching and Learning in Medicine,* 14(1), 34-42. http://dx.doi.org/10.1207/S15328015TLM1401_9

Lakhan, S. (2003). Diversification of U.S. medical schools via affirmative action implementation. *BMC Medical Education,* 3:6. http://dx.doi.org/10.1186/1472-6920-3-6

Lumb, A. B., Homer, M. & Miller, A. (2010). Equity in interviews: Do personal characteristics impact on admission interview scores? *Medical Education,* 44(11), 1077-1083. http://dx.doi.org/10.1111/j.1365-2923.2010.03771.x

Lynch, B., MacKenzie, R., Dowell, J., Cleland, J. & Prescott, G. (2009). Does the UKCAT predict Year 1 performance in medical school? *Medical Education,* 43(12), 1203-1209. http://dx.doi.org/10.1111/j.1365-2923.2009.03535.x

McLaughlin, K. (2014). Are we willing to change our impression of first impressions? *Advances in Health Sciences Education,* 19(3), 429-431. http://dx.doi.org/10.1007/s10459-013-9490-4

McManus, I. C., Powis, D. A., Wakeford, R., Ferguson, E., James, D. & Richards, P. (2005). Intellectual aptitude tests and a levels for selecting UK school leaver entrants for medical school. *BMJ,* 331(7516), 555-559. http://dx.doi.org/10.1136/bmj.331.7516.555

Mercer, A. (2007). *Selecting medical students: An Australian case study.* PhD thesis, Murdoch University. http://researchrepository.murdoch.edu.au/748/

Mercer, A., Abbott, P. V. & Puddey, I. (2012). Relationship of selection criteria to subsequent academic performance in an Australian undergraduate dental school. *European Journal of Dental Education*, 17(1), 39-45. http://dx.doi.org/10.1111/eje.12005

Mercer, A. & Puddey, I. B. (2011). Admission selection criteria as predictors of outcomes in an undergraduate medical course: A prospective study. *Medical Teacher*, 33(12), 997-1004. http://dx.doi.org/10.3109/0142159X.2011.577123

Monash University (2014). Bachelor of Medicine and Bachelor of Surgery (Honours) Admissions information for local applicants 2015 entry. Monash University. http://www.med.monash.edu.au/medical/central/docs/2015-mbbs-domestic-brochure.pdf

O'Neill, L. D., Wallstedt, B., Eika, B. & Hartvigsen, J. (2011). Factors associated with dropout in medical education: A literature review. *Medical Education,* 45(5), 440-454. http://dx.doi.org/10.1111/j.1365-2923.2010.03898.x

Pau, A., Jeevaratnam, K., Chen, Y. S., Fall, A. A., Khoo, C. & Nadarajah, V. D. (2013). The multiple mini-interview (MMI) for student selection in health professions training - a systematic review. *Medical Teacher,* 35(12), 1027-1041. http://dx.doi.org/10.3109/0142159X.2013.829912

Poole, P., Bourke, D. & Shulruf, B. (2010). Increasing medical student interest in general practice in new zealand: Where to from here? *New Zealand Medical Journal,* 123(1315). http://www.nzma.org.nz/journal/read-the-journal/all-issues/2010-2019/2010/vol-123-no-1315/article-poole

Poole, P. & Shulruf, B. (2013). Shaping the future medical workforce: Take care with selection tools. *Journal of Primary Health Care,* 5(4), 269-275. http://www.publish.csiro.au/?act=view_file&file_id=HC13269.pdf

Poole, P., Shulruf, B., Harley, B., Monigatti, J., Barrow, M., Reid, P., Prendergast, C. & Bagg, W. (2012). Shedding light on the decision to retain an interview for medical student selection. *New Zealand Medical Journal,* 125(1361), 81-88. https://www.nzma.org.nz/journal/read-the-journal/all-issues/2010-2019/2012/vol-125-no-1361/view-poole

Poole, P., Shulruf, B., Rudland, J. & Wilkinson, T. (2012). Comparison of UMAT scores and GPA in prediction of performance in medical school: A national study. *Medical Education,* 46(2), 163-171. http://dx.doi.org/10.1111/j.1365-2923.2011.04078.x

Prideaux, D., Roberts, C., Eva, K., Centeno, A., McCrorie, P., McManus, C., Patterson, F., Powis, D., Tekian, A. & Wilkinson, D. (2011). Assessment for selection for the health care professions and specialty training: Consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher,* 33(3), 215-223. http://dx.doi.org/10.3109/0142159X.2011.551560

Puddey, I. & Mercer, A. (2013). Socio-economic predictors of performance in the undergraduate medicine and health sciences admission test (UMAT). *BMC Medical Education,* 13:155. http://dx.doi.org/10.1186/1472-6920-13-155

Puddey, I., Mercer, A., Carr, S. & Louden, W. (2011). Potential influence of selection criteria on the demographic composition of students in an Australian medical school. *BMC Medical Education,* 11:97. http://dx.doi.org/10.1186/1472-6920-11-97

Reiter, H. I., Eva, K. W., Rosenfeld, J. & Norman, G. R. (2007). Multiple mini-interviews predict clerkship and licensing examination performance. *Medical Education,* 41(4), 378-384. http://dx.doi.org/10.1111/j.1365-2929.2007.02709.x

Salvatori, P. (2001). Reliability and validity of admissions tools used to select students for the health professions. *Advances in Health Sciences Education,* 6(2), 159-175. http://dx.doi.org/10.1023/A:1011489618208

Shulruf, B., Poole, P., Wang, G. Y., Rudland, J. & Wilkinson, T. (2012). How well do selection tools predict performance later in a medical program? *Advances in Health Sciences Education,* 17(5), 615-626. http://dx.doi.org/10.1007/s10459-011-9324-1

Silver, B. & Hodgson, C. S. (1997). Evaluating GPAs and MCAT scores as predictors of NBME I and clerkship performances based on students' data from one undergraduate institution. *Academic Medicine,* 72(5), 394-396. http://pdfs.journals.lww.com/academicmedicine/1997/05000/Evaluating_GPAs_and_MCAT_scores_as_predictors_of.22.pdf

Simpson, P., Scicluna, H., Jones, P., Cole, A., O'Sullivan, A., Harris, P., Velan, G. & McNeil, P. (2014). Predictive validity of a new integrated selection process for medical school admission. *BMC Medical Education,* 14:86. http://dx.doi.org/10.1186/1472-6920-14-86

Smetana, G., Landon, B., Bindman, A., Burstin, H., Davis, R., Tjia, J. & Rich, E. (2007). A comparison of outcomes resulting from generalist vs specialist care for a single discrete medical condition: A systematic review and methodologic critique. *Archives of Internal Medicine,* 167(1), 10-20. http://dx.doi.org/10.1001/archinte.167.1.10

University of New South Wales (2012). Selection criteria - local applicants. [viewed 6 Aug 2014]. https://med.unsw.edu.au/selection-criteria-local-applicants

University of Western Australia (2014). Domestic school leaver pathways to MD and DMD. [viewed 1 Sep 2014] http://www.meddent.uwa.edu.au/courses/postgraduate/apply-professional/domestic-school-path

Wilkinson, D., Casey, M. G. & Eley, D. S. (2014). Removing the interview for medical school selection is associated with gender bias among enrolled students. *The Medical Journal of Australia,* 200(2), 96-99. http://dx.doi.org/10.5694/mja13.10103

Wilkinson, D., Zhang, J., Byrne, G. R., Luke, H., Ozolins, I., Parker, M. E. & Peterson, R. (2008). Medical school selection criteria and the prediction of academic performance. *The Medical Journal of Australia,* 188(6), 349-354. https://www.mja.com.au/journal/2008/188/6/medical-school-selection-criteria-and-prediction-academic-performance

Wilkinson, D., Zhang, J. & Parker, M. (2011). Predictive validity of the undergraduate medicine and health sciences admission test for medical students' academic performance. *The Medical Journal of Australia,* 194(7), 341-344. https://www.mja.com.au/journal/2011/194/7/predictive-validity-undergraduate-medicine-and-health-sciences-admission-test

Wood, T. J. (2014). Exploring the role of first impressions in rater-based assessments. *Advances in Health Sciences Education,* 19(3), 409-427. http://dx.doi.org/10.1007/s10459-013-9453-9

Wright, S. R. & Bradley, P. M. (2010). Has the UK clinical aptitude test improved medical student selection? *Medical Education,* 44(11), 1069-1076. http://dx.doi.org/10.1111/j.1365-2923.2010.03792.x

Yeung, A. S., Li, B., Wilson, I. & Craven, R. G. (2013). The role of self-concept in medical education. *Journal of Further and Higher Education,* 38(6), 794-812. http://dx.doi.org/10.1080/0309877X.2013.765944

Dr Colleen Ma is a Junior Doctor at St Vincent's Health Australia. She graduated from Sydney Medical School in 2015.
Email: clkm88@gmail.com

Dr Peter Harris is Senior Lecturer in clinical education in the Office of Medical Education at UNSW. His interests are curriculum, clinical teacher development and assessment with a focus on programmatic assessment.
Email: p.harris@unsw.edu.au

Dr Andrew Cole is Conjoint Associate Professor in Rehabilitation & Aged Care in the School of Public Health & Community Medicine at the University of New South Wales, Sydney, Australia, and Chief Medical Officer of HammondCare. Andrew's research interest is in development of healthcare and aged care services and their staff, with particular interests in student and staff selection, interdisciplinary training, and measurement of educational outcomes, and relating these to improving the provision of service to older and disabled people in need of care.
Email: acole@unsw.edu.au

Professor Philip Jones was the Deputy Dean Education in the Faculty of Medicine, University of New South Wales, Sydney, Australia. His research interest is in educational assessment with a particular focus on the assessment of clinical competence. He is currently the Senior Assessment Consultant in the Office of the Deputy Vice Chancellor Education.
Email: Philip.jones@unsw.edu.au

Dr Boaz Shulruf (corresponding author) is Associate Professor in Medical Education Research at University of New South Wales, Sydney, Australia. Boaz's research interest is in educational assessment, with particular focus upon topics related to setting of standard setting, student selection, and measurement of educational outcomes.
Email: b.shulruf@unsw.edu.au

Please cite as: Ma, C., Harris, P., Cole, A., Jones, P. & Shulruf, B. (2016). Selection into medicine using interviews and other measures: Much remains to be learned. *Issues in Educational Research*, 26(4), 623-634. http://www.iier.org.au/iier26/ma.pdf