מרכז ארצי לבחינות ולהערכה (ע"ר)
NATIONAL INSTITUTE FOR TESTING & EVALUATION
المركز القطري للامتحانات والتّقييم
מיסודן של האוניברסיטאות בישראל

The Predictive Validity of a Two-Step Selection Process to Medical Schools

Report

RR 16-03

March 2016

Tamar Kennet-Cohen | Elliot Turvall | Yonatan Saar | Carmel Oren

NITE REPORT RR-16-03

ISBN:978-965-502-195-0

Table of Contents

Abstract

Background: A two-step selection process, consisting of cognitive and non-cognitive measures, is common in medical school admissions.

Aim: To estimate the validity of this process in predicting academic performance, taking into account the complex and pervasive effect of range restriction in this context.

Method: The estimation of the validity of the two-step process included a sequential correction for range restriction and an estimation of the predictive validity of the process in its entirety. Data were collected from 1,002 undergraduate students from four cohorts (2006/07-2009/10) at three medical schools in Israel.

Results: The predictive validity of the composite of the cognitive measures with respect to Year 1 performance was high, resulting entirely from the predictive validity of the admission test (a standard measure of ability). The predictive validity of the non-cognitive measure was moderate. The predictive validity of the process in its entirety was high, its value dependent on the weights given to the cognitive and non-cognitive measures.

Conclusion: A cognitive admission test has a high predictive validity with respect to Year 1 performance. The addition of a non-cognitive measure in the second step does not markedly diminish the predictive validity of the selection process with respect to academic achievement.

Introduction

Predictive validity studies for medical school admissions

Medical schools invest considerable resources in designing selection processes that identify the most appropriate candidates for their programs. Given the high-stakes nature of these processes, it is essential to establish their predictive validity.

Most medical school admissions offices rely on a combination of cognitive and non-cognitive measures. The cognitive measures typically consist of an admission test and a measure of prior academic achievement. In North American (graduate-entry) medical schools, these include the Medical College Admission Test (MCAT) and undergraduate grade point average (uGPA). The predictive validity of MCAT is well established: Recent meta-analytic studies show that it has a predictive value for pre-clinical performance, with correlations (corrected for range restriction) of 0.43-0.64 with GPAs in the first two years of medical school (Julian 2005; Donnon et al. 2007; Kreiter & Krieter 2007). Results with respect to uGPA are also acceptable, though somewhat lower (Julian 2005; Kreiter & Krieter 2007). The critical and consistent finding, however, is that MCAT scores and uGPAs together predict performance better than either considered alone.

Data regarding the predictive validity of cognitive measures outside of North America are relatively scarce (Prideaux et al. 2011). Notable examples include the Graduate Medical School Admissions Test (GAMSAT) given in Australia and the United Kingdom, and the Undergraduate Medicine and Health Sciences Admission Test (UMAT) and UK Clinical Aptitude Test (UKCAT), given in Australia and the UK, respectively. These last two tests assess general cognition, unlike the MCAT or the GAMSAT, which examine candidates' knowledge of basic medical science. Studies of the GAMSAT (Coates 2008), UMAT (Mercer & Puddey 2011; Wilkinson et al. 2011; Poole et al. 2012; Edwards et al. 2013) and UKCAT (McManus et al. 2013) show that their predictive validity ranges from low to moderate with respect to medical school performance in the first years. The predictive validity of prior academic achievement is generally higher. As in the North American context, most of these studies found that a combination of admission test and prior academic achievement provided the best means of predicting Year 1 performance.

The above literature review concentrates on the prediction of pre-clinical performance in medical school. This criterion, adopted in the current study, reflects an individual's capability to successfully undertake a high-level, demanding curriculum. As for the prediction of clinical performance, recent reviews concluded that MCAT and uGPA have a positive relationship with clinical skills, although this relationship is weaker than that with cognitive outcome variables (Kreiter & Krieter 2007; Siu & Reiter 2009).

Non-cognitive measures used in medical school admissions include traditional tools such as interviews, personal statements, autobiographical statements or letters of recommendation, and relatively novel tools, such as the multiple mini-interview (MMI) and situational judgment test (SJT) (Eva et al. 2004; Lievens & Sackett 2006). Reviews summarizing the current state of findings on the predictive validity of medical school selection measures conclude that there is no evidence regarding the predictive validity of traditional non-cognitive measures (Siu & Reiter 2009; Prideaux et al. 2011; Kreiter & Axelson 2013).

As for novel non-cognitive measures, studies of MMI have indicated good predictive validity with respect to performance criteria from the early stages of medical school through licensing examinations (Eva et al. 2004; Reiter et al. 2007; Eva et al. 2009; Eva et al. 2012; Husbands & Dowell 2013). As expected, predictive validity vis-à-vis assessments that focus on clinical aspects of medicine is higher. Nevertheless, it also predicts assessments that are more "cognitive" and knowledge oriented. The SJT has emerged as a potential predictor of academic performance in interpersonal skills courses taken during the first years of medical study and later, of performance in actual interpersonal situations, such as internships and jobs (Lievens et al. 2005; Lievens & Sackett 2012). Its correlation with academic performance is weaker but still significant (Lievens & Sackett 2006).

Thus, there is an apparent complementary relationship between cognitive and non-cognitive measures, indicating a clear advantage to combining them in the selection process. Such a combination is usually achieved through a two-step procedure, whereby initial selection and final selection are based on cognitive and non-cognitive predictors, respectively. This paper focuses on estimating the predictive validity of such a process.

The admissions process at medical schools in Israel

This study examines the predictive validity of the admissions process, combining cognitive and non-cognitive measures, used by three undergraduate-entry medical schools in Israel. The cognitive measures consist of the Psychometric Entrance Test (PET) and high school matriculation certificate (Bagrut).

The non-cognitive measure is an assessment center, either the MOR (Hebrew acronym for 'selection for medicine') or MIRKAM (Hebrew acronym for 'system of short and structured interviews'), both inspired by the MMI model. MOR is used in two medical schools, since 2004 and 2006, respectively, and MIRKAM is used in one medical school, since 2006.

The medical school admissions process combines these cognitive and non-cognitive measures in two steps: (1) Candidates are rank-ordered on the basis of a composite score comprised of the two cognitive measures (Composite1). Candidates with the highest Composite1 score are invited to take MOR or MIRKAM. (2) A second composite score, comprised of the Composite1 score and the score on MOR or MIRKAM (Composite2), is used for the final admissions decision. The weights attached to the components of Composite1 and Composite2 differ between medical schools and cohorts.

Aims

This paper's contribution is threefold: (1) The study contributes to the small body of work on the validity of selection processes used by medical schools outside of North America. (2) Given the extreme selection ratio that typically characterizes the selection process to medical schools, the values of the observed (uncorrected) validity coefficients are generally misleading. Yet, it is not uncommon to find predictive validity studies refraining from applying correction for range restriction simply because the selection process is so complex. This study describes correction for range restriction in the complex, yet prevalent, context of a two-step selection procedure (Sackett & Yang 2000). (3) There is a tendency in the literature to focus on the properties of individual selection measures (Prideaux et al. 2011), but ultimately, the validity of the decision procedure as a whole, rather than the validity of its individual selection measures, is what matters. In this spirit, we present a simulation-based estimation of the predictive validity of the entire process.

Method

Participants

Participants in the study included 1,002 first-year students from four cohorts (2006/07-2009/10) at the Tel Aviv University, The Technion - Israel Institute of Technology and Hebrew University of Jerusalem undergraduate-entry medical schools.

Study variables

The criterion measure was Year 1 grade point average (GPA) at medical school. The cognitive predictors were Bagrut and the PET. The Bagrut score is an average of the scores in various high school subjects, weighted by the scope of studies in each subject. The score in a single subject is computed as an average of the school grade and the score obtained on a national test. The PET is a general scholastic aptitude test. It includes three multiple-choice subtests: Verbal Reasoning, Quantitative Reasoning and English as a foreign language. (In 2012, a writing task was added to the Verbal Reasoning subtest but this is irrelevant to the study cohorts.) The PET total score is computed as the average of scores on the Verbal Reasoning, Quantitative Reasoning and English subtests, weighted 40/40/20, respectively (Beller 1994; Kennet-Cohen et al. 2014; Oren et al. 2014).

The non-cognitive predictors were MOR and MIRKAM. They include nine or eight behavioral stations (BS), respectively, and two customized questionnaires – the Judgment and Decision-Making Questionnaire (JDQ) and a standard Biographical Questionnaire (BQ). The nature of the two systems' behavioral stations is different and has been described elsewhere (Gafni et al. 2012; Ziv et al. 2012). MOR and MIRKAM total scores are computed as the average of scores on the JDQ, BQ and BS, weighted 20/20/60, respectively, for MOR and 15/25/60, respectively, for MIRKAM. MOR and MIRKAM are not identical but are designed to measure the same personal attributes.

On the basis of these predictors, we computed the combination of the PET score and Bagrut score (both standardized) in equal weights (Composite1) and the combination of the Composite1 score and the MOR or MIRKAM score (also standardized) in equal weights (Composite2-MOR and Composite2-MIRKAM, respectively) as three additional predictors.

Statistical analyses

Pearson's correlations between Year 1 GPA and the predictors were computed separately for each combination of medical school and cohort (these 12 combinations are referred to hereafter as "academic units"). Correction for range restriction for the two-step selection was conducted (separately for each academic unit) by repeatedly applying the correction formula as follows (Sackett & Yang 2000): first, we treated the admitted students as the restricted group and the candidates invited to take the MOR or MIRKAM as the unrestricted group, and applied a correction formula. Then we treated the candidates invited to take the MOR or MIRKAM as the restricted group and all the applicants as the unrestricted group, and a correction formula was applied to the previously corrected correlations.

The correction formula was suitable for incidental selection (selection on a third variable) (Gulliksen 1950). To apply this formula, it is necessary to identify the variable on which selection was based. That variable – similar to Composite1 in step 1 and Composite2 in step 2 – differed between medical schools and cohorts in the weights attached to their components. The weights attached to the components of the variable used for selection in stage 1 ranged from 50% for the PET and 50% for the Bagrut to 70% for the PET and 30% for the Bagrut, and the weights attached to the components of the variable used for selection in stage 2 ranged from 50% for the variable used in stage 1 and 50% for the MOR or the MIRKAM to 0% for the variable used in stage 1 and 100% for the MOR or the MIRKAM. In applying the correction formula to each stage, we used the specific selection variable suitable for each academic unit.

Next, we conducted Monte Carlo simulations to examine the predictive validity of two prototype two-step selection rules. The first step in both of them was based on Composite1. The second step was based on the MOR in one selection rule and on Composite2-MOR in the other. These two-step selection rules were compared to three one-step selection rules based on Composite1, the MOR or Composite2-MOR. A pool of 1,000 applicants was used for each simulation, with variables defined as having multivariate N(0,1) distribution. Correlations between the variables were set according to the empirical data. Given the context of medical school selection, the group proceeding to the second step (in the two-step selection rules) consisted of 400 applicants, and the admitted group 100 applicants. The simulation of each of the five selection rules described above was replicated 500 times and the results presented are averages across the 500 replications.

Results

Predictive validity of the admission measures

Table 1 displays the correlations between the measures of the selection process and Year 1 GPA. Observed coefficients as well as coefficients corrected for range restriction are presented. The results are summarized across the 12 academic units by reporting the weighted mean and interquartile range of the correlation coefficients.

Table 1 – Correlations between measures of the selection process and Year 1 GPA
(values for observed correlations in brackets)

Predictor	Mean		Interquartile range	
Bagrut	0.30*	(0.05*)	0.14 - 0.56	(-0.12 - 0.25)
PET total	0.61***	(0.15***)	0.49 - 0.84	(0.07 - 0.27)
PET V	0.55***	(0.01)	0.41 - 0.78	(-0.10 - 0.11)
PET Q	0.57***	(0.14***)	0.42 - 0.73	(0.07 - 0.24)
PET E	0.54***	(0.14***)	0.48 - 0.79	(0.13 - 0.18)
Composite1	0.61***	(0.18***)	0.49 - 0.82	(-0.02 - 0.33)
MOR total	0.35***	(0.17***)	0.04 - 0.70	(0.15 - 0.26)
MIRKAM total	0.37**	(0.09**)	0.13 - 0.53	(0.03 - 0.17)
JDQ	0.31**	(0.09**)	0.14 - 0.44	(0.01 - 0.20)
BQ	0.33**	(0.10***)	0.20 - 0.56	(-0.02 - 0.26)
MOR BS	0.30***	(0.14***)	0.05 - 0.47	(0.09 - 0.24)
MIRKAM BS	0.27**	(0.05)	0.06 - 0.42	(-0.01 - 0.13)
Composite2-MOR	0.63***	(0.28***)	0.53 - 0.86	(0.17 - 0.36)
Composite2- MIRKAM	0.64***	(0.24***)	0.51 - 0.86	(0.12 - 0.38)

Bagrut = high school matriculation certificate score; from PET total through PET E = the Psychometric Entrance Test total and section (Verbal reasoning, Quantitative reasoning and English) scores, respectively; Composite1 = the combination of the PET total score and the Bagrut score in equal weights; MOR total and MIRKAM total = the total score in MOR and MIRKAM assessment centers; JDQ and BQ = Judgement and Decision-making Questionnaires and Biographical Questionnaires, both being a part of MOR and MIRKAM; MOR BS and MIRKAM BS = behavioral stations in MOR and MIRKAM, respectively; Composite2-MOR and Composite2-MIRKAM = the combination of Composite1 score and the MOR or the MIRKAM total score, respectively, in equal weights.

**p* <0.05; ***p* <0.01; ****p* <0.001. *p*-values were computed for the means of both the observed and the corrected correlations, based on the standard errors of the individual correlations (Borenstein at al. 2009). The standard errors of the individual corrected correlations were computed by extending the formula for direct selection (Bobko & Rieck 1980) to the case of indirect selection and applying it sequentially for the two-stage selection.

n=1,002 except for MOR BS, MOR total and Composite2-MOR (*n*=902) and MIRKAM BS, MIRKAM total and Composite2-MIRKAM (*n*=879). The score in MOR is required by two of the medical schools and the score in MIRKAM by the third. However, many applicants (*n*=779) have scores in both systems, which enabled us to examine their validities in all the medical schools.

The predictive validities of the PET subtests are uniformly high, ranging from 0.54 to 0.57 (corrected). The predictive validity of the PET total score is 0.61, revealing that the combination of the subtest scores is a more powerful predictor of Year 1 GPA than is any subtest alone. The predictive validity of the Bagrut is 0.30. The predictive validity of Composite1 is 0.61, identical to that of the PET total score.

The predictive validities of the MOR and MIRKAM components range from 0.27 to 0.33 (corrected). The predictive validities of MOR and MIRKAM total scores are 0.35 and 0.37, respectively. Combining the cognitive and non-cognitive measures into Composite2-MOR and Composite2-MIRKAM leads to predictive validities of 0.63 and 0.64, marginally higher than the predictive validity of the composite of the cognitive measures alone (0.61).

Finally, the size of the interquartile range shows that there is a noticeable variation in the size of the correlations among the 12 academic units, implying that the admission measures are better predictors of performance in certain contexts than in others.

Predictive validity of the two-step selection process

The validity coefficients reported in Table 1 convey information relevant to a course of events in which the predictors are used in a single-step selection. Table 2 presents the results of two simulated two-step selection rules. In both of them, the first step is based on Composite1. The second step is based either on the MOR or Composite2-MOR. These two-step selection rules were compared with one-step selection rules based on their constituent predictors.

Table 2 – Results of several simulated selection rules

	Predictors	**Mean of Year 1 score**	**Correlation with Year 1 score**
One-step selection rules	Composite1	1.07	0.61
	MOR	0.61	0.35
	Composite2-MOR	1.11	0.63
Two-step selection rules	1st step: Composite1, 2nd step: MOR	0.93	0.53
	1st step: Composite1, 2nd step: Composite2-MOR	1.11	0.63

Composite1 = a hypothetical variable simulating the combination of the PET total score and the Bagrut score in equal weights; MOR = a hypothetical variable simulating the total score in MOR assessment center; Composite2-MOR = a hypothetical variable simulating the combination of Composite1 score and the MOR total score in equal weights.

The results of the different selection rules are presented in terms of the mean Year 1 score among the admitted students, bearing in mind that the Year 1 score is defined as a standard normal variable among applicants. In addition, the predictive validity of the selection rules are presented in terms of the correlation with Year 1 score: for the one-step rules, these values are identical to the respective correlations presented in Table 1; for the two-step rules, these values were derived in a simulation of a one-step selection rule leading to the designated mean Year 1 score. Table 2 shows that a transition from a one-step selection rule based on Composite1 to a two-step selection rule based on Composite1 in the first step and on MOR in the second step lowers the mean Year 1 score among admitted students. However, it still leads to a gap of almost a full standard deviation between the mean Year 1 score among applicants (0) and among admitted students (0.93). Such a gap between applicants and admitted students would be obtained from a one-step selection rule using a predictor whose correlation with the Year 1 score is 0.53. When the two-step selection rule is based on the Composite2-MOR in the second step, its predictive validity (0.63) is similar to the predictive validity of the one-step selection rule based on Composite1 (0.61).

Discussion

The findings support the predictive validity of a selection process that combines cognitive and non-cognitive measures in predicting academic performance.

As for the cognitive measures, the study is a confirmation that a standard objective aptitude test, the PET, is the most valid aspect of the selection process. Its standardization allows comparability between applicants from diverse secondary schools and different cohorts, which the Bagrut does not do. In fact, the results show that PET scores predict Year 1 GPA as accurately as the two cognitive measures together. It should be noted that the PET, like the UMAT and the UKCAT (and unlike the MCAT or the GAMSAT), is a test of general cognitive skills. However, unlike the UMAT and the UKCAT, which are designed especially to measure aptitude for medical study, the PET is used in admissions to most undergraduate programs in Israeli universities, and none of its components is specifically designed for medical school admissions. Thus, our findings regarding the PET's predictive validity can add unique evidence regarding cognitive predictors of medical school performance.

As for the non-cognitive measures, the correlations between MOR and MIRKAM and academic performance are not expected to be high, but they are not expected to be null either. There is a general consensus that MMIs are not devoid of cognitive components (Prideaux et al. 2011), thus it is not surprising that MOR and MIRKAM have a medium correlation with Year 1 GPA. High performance on these measures depends on the ability to comprehend social situations, capture complex and problematic situations, and express oneself both orally and in writing. Furthermore, Year 1 performance does predominantly reflect factual knowledge, but subject matter is not limited to the biological and physical sciences. Each of the three medical schools participating in our study offers Year 1 courses (e.g., "medical education and communication", "medical ethics", "humane aspects of medicine") that deal with interpersonal communication, medical ethics, cultural competence, professionalism, etc. The proportion of these courses in the Year 1 program is not high (up to 13%) but nonetheless, they may contribute to the predictive validity of the non-cognitive measures (Lievens & Sackett 2006).

Aside from establishing the validity of the individual measures used in the selection process, this study presented an estimation of the selection process as a whole, which revealed that including an assessment of non-cognitive qualities in the selection process does not necessarily mean that the predictive validity of the process with respect to academic achievement is substantially compromised. A similar finding was reported elsewhere (Terregino et al. 2015).

The results highlight large differences in the correlations obtained for different academic units. Such variation in the strength of correlations between cohorts and institutions is a typical finding (Julian 2005; Coates 2008; Edwards et al. 2013), emphasizing the pitfalls of generalizing results from predictive validity studies without recognizing differences in the contexts in which they are conducted. Such a variation can be viewed as a form of sampling error, which leads to our ultimate conclusion that the ongoing effort to accumulate data from numerous contexts is vital to establishing a sound understanding of what works in the process of medical school admissions. Thus, while the results reported here are not necessarily generalizable to every specific context of medical school admissions, they relate to a widespread procedure of combining cognitive and non-cognitive measures in a multi-step procedure. In this way, they contribute to the cumulative knowledge regarding the predictive validity of medical school admission procedures, with an eye towards those outside of North America.

This study had several limitations. Firstly, corrections for range restriction are based on several assumptions. The corrected correlations may be biased due to violations of these assumptions, but nonetheless, they are generally more accurate than the uncorrected correlations (Linn 1983). Similarly, our simulation of the admissions situation was based on an assumption of multivariate normality. While an alternative assumption of multivariate uniform distribution yielded identical results, we should remember that the results are hypothetical and should be viewed as approximations. Thirdly, under the meritocratic approach which characterizes medical school admissions decisions, predictive validity is the most important feature of selection procedures (Trost 1995). Clearly, other criteria for evaluating selection procedures, such as selection bias and indirect effects, are outside the scope of the present study, and deserve consideration in future research. Finally, this study was restricted to first-year performance. Previous results show that undergraduate and post-graduate performance across years is highly stable, with first-year medical school performance strongly predicting subsequent performance (McManus et al. 2013). Still, it will be important to extend this study to later years. The predictive validity of the selection process for students in the clinical stage of training will be the focus of a future study when sufficient data for our cohorts are available.

Acknowledgments: The authors wish to thank Dvir Kleper, National Institute for Testing and Evaluation, for his methodological contributions, and Marlene Korenstein, National Institute for Testing and Evaluation, for copyediting.

References

Beller M. 1994. Psychometric and social issues in admissions to Israeli universities. Educational Measurement: Issues and Practice 13:12–20.

Bobko P, Rieck A. 1980. Large sample estimators for standard errors of functions of correlation coefficients. Appl Psych Meas 4:385–398.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. 2009. Introduction to Meta-Analysis. Hoboken, NJ: Wiley.

Coates H. 2008. Establishing the criterion validity of the Graduate Medical School Admissions Test (GAMSAT). Med Educ 42:999–1006.

Donnon T, Oddone-Paolucci EO, Violato C. 2007. The predictive validity of the MCAT for medical school performance and medical board licensing examinations: A meta-analysis of the published research. Acad Med 82:100–106.

Edwards D, Friedman T, Pearce J. 2013. Same admissions tools, different outcomes: A critical perspective on predictive validity in three undergraduate medical schools. BMC Medical Education 13:173.

Eva KW, Reiter HI, Rosenfeld J, Norman GR. 2004. The ability of the multiple mini-interview to predict preclerkship performance in medical school. Acad Med 79(S10):S40–42.

Eva KW, Reiter HI, Rosenfeld J, Trinh K, Wood TJ, Norman GR. 2012. Association between a medical school admission process using the multiple mini-interview and national licensing examination scores. JAMA 308:2233–2240.

Eva KW, Reiter HI, Trinh K, Wasi P, Rosenfeld J, Norman GR. 2009. Predictive validity of the multiple mini-interview for selecting medical trainees. Med Educ 43:767–775.

Gafni N, Moshinsky A, Eisenberg O, Zeigler D, Ziv A. 2012. Reliability estimates: Behavioral stations and questionnaires in medical school admissions. Med Educ 46:277–288.

Gulliksen H. 1950. Theory of mental tests. New York: Wiley.

Husbands A, Dowell J. 2013. Predictive validity of the Dundee multiple mini-interview. Med Educ 47:717–725.

Julian ER. 2005. Validity of the Medical College Admission Test for predicting medical school performance. Acad Med 80:910–917.

Kennet-Cohen T, Turvall E, Oren C. 2014. Detecting bias in selection for higher education: Three different methods. Assessment in Education: Principles, Policy & Practice 21:193–204.

Kreiter CD, Axelson RD. 2013. A perspective on medical school admission research and practice over the last 25 years. Teach Learn Med 25(S1):S50–56.

Kreiter CD, Kreiter Y. 2007. A validity generalization perspective on the ability of undergraduate GPA and the Medical College Admission Test to predict important outcomes. Teaching and Learning in Medicine 19:95–100.

Lievens F, Buyse T, Sackett PR. 2005. The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. J Appl Psychol 90:442–452.

Lievens F, Sackett PR. 2006. Video-based vs. written situational judgment tests: A comparison in terms of predictive validity. J Appl Psychol 91:1181–1188.

Lievens F, Sackett PR. 2012. The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. J Appl Psychol 97:460–468.

Linn R. 1983. Pearson selection formulas: Implications for studies of predictive bias and estimates of educational effects in selected samples. J Educ Meas 20:1–15.

McManus IC, Dewberry C, Nicholson S, Dowell J. 2013. The UKCAT-12 study: Educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a cross-sectional collaborative study of twelve UK medical schools. BMC Med 11:244.

McManus IC, Woolf K, Dacre J, Paice E, Dewberry C. 2013. The academic backbone: Longitudinal continuities in educational achievement from secondary school and medical school to MRCP(UK) and the Specialist Register in UK medical students and doctors. *BMC Medicine* 11:242.

Mercer A, Puddey IB. 2011. Admission selection criteria as predictors of outcomes in an undergraduate medical course: A prospective study. Med Teach 33:997–1004.

Oren C, Kennet-Cohen T, Turvall E, Allalouf A. 2014. Demonstrating the validity of three general scores of PET in predicting higher education achievement in Israel. Psicothema 26:117–126.

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. Med Educ 46:163–171.

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 for Ottawa 2010 Conference. Med Teach 33:215–223.

Reiter HI, Eva KW, Rosenfeld J, Norman GR. 2007. Multiple mini-interviews predict clerkship and licensing examination performance. Med Educ 41:378–384.

Sackett PR, Yang H. 2000. Correction for range restriction: An expanded typology. J Appl Psychol 85:112–118.

Siu E, Reiter HI. 2009. Overview: What's worked and what hasn't as a guide towards predictive admissions tool development. Adv Health Sci Ed 14:759–775.

Terregino CA, McConnell M, Reiter HI. 2015. The effect of differential weighting of academics, experiences, and competencies measured by Multiple Mini Interview (MMI) on race and ethnicity of cohorts accepted to one medical school. Acad Med 90:1651–1657.

Wilkinson D, Zhang J, Parker M. 2011. Predictive validity of the undergraduate medicine and health sciences admission test for medical students' academic performance. MJA 194:341–344.

Ziv A, Rubin O, Moshinsky A, Gafni N, Kotler M, Dagan Y, Lichtenberg D, Mekori Y, Mittleman M. 2012. MOR: A simulation-based assessment center for evaluating the personal and interpersonal qualities of medical school candidates. Med Educ 42:991–998.

Trost G. 1995. Principles and practice in selection for admission to higher education. In: Kelliaghan, ed. Admission to high education: Issues and practice. Princeton, NJ: International Association for Educational Assessment, pp. 7–15.