Research Article

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

Tamar Kennet-Cohen, Elliot Turvall, Yonatan Saar, and Carmel Oren

National Institute for Testing and Evaluation, P.O.B. 26015, 91260 Jerusalem, Israel

Correspondence should be addressed to Tamar Kennet-Cohen; tami@nite.org.il

Background. A two-step selection process, consisting of cognitive and noncognitive measures, is common in medical school admissions. *Objective.* 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. *Methods.* 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 noncognitive 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 noncognitive measures. *Conclusion.* A cognitive admission test has a high predictive validity with respect to Year 1 performance. The addition of a noncognitive measure in the second step does not markedly diminish the predictive validity of the selection process with respect to academic achievement.

1. Introduction

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 admissions offices rely on a combination of cognitive and noncognitive measures. The former 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 preclinical performance, with correlations (corrected for range restriction) of 0.43–0.64 with GPAs in the first two years of medical school [1–3]. Results with respect to uGPA are also acceptable, though somewhat lower [2, 3].

Data regarding the predictive validity of cognitive measures outside of North America are relatively scarce [4].

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. Studies of the GAMSAT [5], UMAT [6–9], and UKCAT [10] 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.

The above literature review concentrates on the prediction of preclinical 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 predicting 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 [3, 11].

Noncognitive 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 tests (SJT) [12, 13]. 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 noncognitive measures [4, 11, 14]. As for novel noncognitive measures, studies of MMI have indicated good predictive validity with respect to performance criteria from the early stages of medical school through licensing examinations [12, 15–18]. 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 [19, 20]. Its correlation with academic performance was weaker but still significant [13].

Thus, there is an apparent complementary relationship between cognitive and noncognitive measures, indicating a clear advantage in 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 noncognitive predictors, respectively.

This study focuses on estimating the predictive validity of such a two-step process, combining cognitive and noncognitive measures, which is 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 noncognitive 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.

These cognitive and noncognitive measures are combined in two steps: (1) candidates are rank-ordered on the basis of a composite score comprised of the two cognitive measures (Composite 1). Candidates with the highest Composite 1 score are invited to take MOR or MIRKAM. (2) A second composite score (Composite 2), comprised of Composite 1 score and the score on MOR or MIRKAM, is used for the final admissions decision.

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 medical school admissions, the values of the observed (uncorrected) validity coefficients are generally misleading. This study describes correction for range restriction in the complex, yet prevalent, context of a two-step selection procedure [21]. (3) There is a tendency in the literature to focus on the properties of individual selection measures [4], 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.

2. Methods

2.1. Participants. Participants 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 the Hebrew University of Jerusalem undergraduate-entry medical schools.

2.2. Study Variables. The criterion measure was medical school Year 1 grade point average (GPA).

The cognitive predictors were the 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) [22].

The noncognitive predictors were MOR and MIRKAM. MOR is used by two of the three medical schools (since 2004) and MIRKAM is used by the third (since 2006). MOR and MIRKAM 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 questionnaires used by both systems are identical, but the nature of their behavioral stations is different [23, 24]. Nonetheless, MOR and MIRKAM are designed to measure the same personal attributes.

On the basis of these predictors, we computed the combination of the PET score and the Bagrut score in equal weights (Composite 1) and the combination of Composite 1 score and the MOR or the MIRKAM score in equal weights (Composite 2-MOR and Composite 2-MIRKAM, resp.) as three additional predictors. In a preliminary study, we examined several schemes for weighting the components of the composite scores. In order to focus on the most significant results, we limit ourselves here to discussing composite scores in which the components are weighted equally.

2.3. 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 recommended for sequential selection [21]: 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 the one suitable for incidental selection [25]. To apply this formula, it is necessary to identify

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)
Composite 1	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)
Composite 2-MOR	0.63** (0.28***)	0.53–0.86 (0.17–0.36)
Composite 2-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 subtest (Verbal Reasoning, Quantitative Reasoning, and English) scores, respectively; Composite 1 = 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 Questionnaire and Biographical Questionnaire, both being part of MOR and MIRKAM; MOR BS and MIRKAM BS = behavioral stations in MOR and MIRKAM, respectively; and Composite 2-MOR and Composite 2-MIRKAM = the combination of Composite 1 score and the MOR or the MIRKAM total score, respectively, in equal weights.
*$p < 0.05$; **$p < 0.01$; and ***$p < 0.001$ after Bonferroni correction for multiple tests. p values were computed for the *means* of both the observed and the corrected correlations, based on the standard errors of the individual correlations [28]. The standard errors of the individual corrected correlations were computed by extending the formula for direct selection [29] to the case of indirect selection and applying it sequentially to the two-stage selection. $n = 1,002$ except for MOR BS, MOR total, and Composite 2-MOR ($n = 902$) and MIRKAM BS, MIRKAM total, and Composite 2-MIRKAM ($n = 879$). As mentioned above, MOR is used by two of the medical schools ($n = 631$) and MIRKAM is used by the third ($n = 371$). However, many applicants are tested in both systems.

the variable on which selection was based. That variable—similar to Composite 1 in step (1) and Composite 2 in step (2)—differed between medical schools and cohorts in the weights attached to their components. In applying the correction formula to each step, we used the specific selection variable suitable for each academic unit.

Next, in order to estimate the predictive validity of the selection process in its entirety, we conducted Monte Carlo simulations of several selection rules. The hypothetical variables we constructed simulated our criterion variable (mean Year 1 score) and three predictors: Composite 1, as a representative of the cognitive predictors; MOR, as a representative of the noncognitive predictors; and Composite 2-MOR, representing a combination of cognitive and noncognitive predictors in equal weights. Two of the selection rules we simulated were two-step rules. The first step in both of them was based on Composite 1, and the second step was based on either the MOR or Composite 2-MOR. The three other selection rules we simulated were one-step rules,

each based on a single predictor—Composite 1, the MOR, or Composite 2-MOR.

A pool of 1,000 hypothetical applicants was used for each simulation, with variables—the criterion and the relevant predictors—defined as having multivariate $N(0, 1)$ distribution. The correlations between the variables were set according to the empirical data. We applied the different selection rules to this pool of applicants. Given the context of medical students' selection, the group proceeding to the second step (in the two-step selection rules) consisted of 400 applicants, and the admitted group consisted of 100 applicants. Following the application of each selection rule, we computed the average score of the criterion variable for the admitted 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.

3. Results

3.1. 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.

The predictive validities of the PET subtests are uniformly high, ranging from 0.54 to 0.57 (corrected). The predictive validities of the PET total score and the Bagrut are 0.61 and 0.30, respectively, the latter being nonsignificant after Bonferroni correction. The predictive validity of Composite 1 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 noncognitive measures into Composite 2-MOR and Composite 2-MIRKAM leads to predictive validities of 0.63 and 0.64, respectively, 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.

3.2. 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 Composite 1. The second step is based either on the MOR or on the Composite 2-MOR. These two-step selection rules were compared with one-step selection rules based on their constituent predictors.

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 Year 1 score is defined as a standard normal variable among applicants. In addition, the predictive

TABLE 2: Results of several simulated selection rules.

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

Composite 1 = 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; and Composite 2-MOR = a hypothetical variable simulating the combination of Composite 1 score and the MOR total score in equal weights.

validity of the selection rules is 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 Composite 1 to a two-step selection rule based on Composite 1 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 scores 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 Year 1 score is 0.53. When the two-step selection rule is based on Composite 2-MOR in the second step, its predictive validity (0.63) is similar to the predictive validity of the one-step selection rule based on Composite 1 (0.61).

4. Discussion

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

As for the cognitive measures, the study is 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. The lower predictive validity of the Bagrut might be explained, at least partially, by the fact that the Israeli education system reflects the diversity of Israeli society: different sectors of the population (e.g., secular, religious, and Arab sectors) attend different schools, which offer different academic curricula. Furthermore, there are separate matriculation examination grading systems for the various sectors, even for the same subject. As a result, the Bagrut score is not fully comparable across sectors. Despite these limitations, as well as a lack of standardization and calibration of the Bagrut subject scores within sectors, the Bagrut validity is generally higher [26] than that reported in our study. In our study, the proportions of students from different sectors are not typical, and, given the complex relationship between Bagrut score, Year 1 GPA, and sector [27], this can explain why our

pattern of results diverges from what we normally see. Indeed, within our study, the Bagrut score was least valid within those academic units where the proportions of students from different sectors were most atypical. As for the PET, it should be noted that, like the UMAT and the UKCAT (and unlike the MCAT or the GAMSAT), the PET 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 for 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 noncognitive 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 [4]; 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 interactions, capture complex and problematic situations, and express oneself both orally and in writing. Furthermore, while Year 1 performance does predominantly reflect factual knowledge, subject matter is not limited to the biological and physical sciences, but it also includes courses focusing on interpersonal skills, where it was found that performance could be predicted by noncognitive measures [13].

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, revealing that including an assessment of noncognitive qualities in the selection process does not necessarily mean that the predictive validity of the process with respect to academic achievement is substantially compromised.

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 [2, 5, 6], 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.

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. A large body of research has shown that they are generally more accurate than the uncorrected correlations [30], but this clearly does not apply to each and every specific study. Therefore, the issue of the accuracy of the corrections should be taken into account by noting that the true correlations may be lower than the ones we reported, but they are clearly positive and have pragmatic use. Secondly, 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 aspect of selection procedures. 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 postgraduate performance across years are highly stable, with first-year medical school performance strongly predicting subsequent performance [31]. 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.

5. Conclusion

A standard objective cognitive admission test has high predictive validity with respect to medical school Year 1 performance. The addition of a noncognitive measure in the second step of the selection process does not markedly diminish the predictive validity of the selection process with respect to academic achievement.

Competing Interests

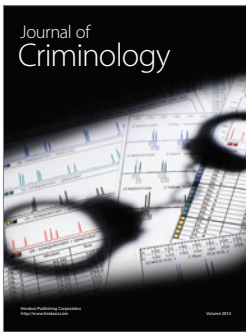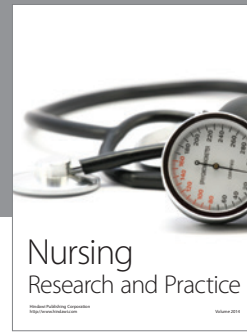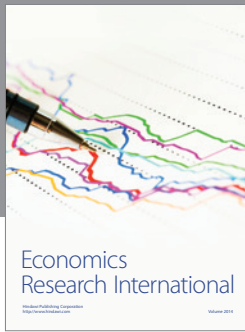The authors declare that there is no conflict of interests regarding the publication of this paper.
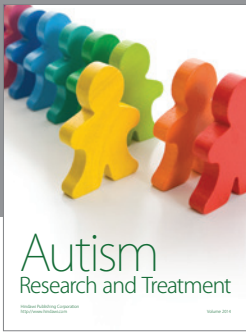
Acknowledgments

References

[1] T. Donnon, E. O. Paolucci, and C. Violato, "The predictive validity of the MCAT for medical school performance and medical board licensing examinations: a meta-analysis of the published research," *Academic Medicine*, vol. 82, no. 1, pp. 100–106, 2007.

[2] E. R. Julian, "Validity of the medical college admission test for predicting medical school performance," *Academic Medicine*, vol. 80, no. 10, pp. 910–917, 2005.

[3] C. D. Kreiter and Y. Kreiter, "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*, vol. 19, no. 2, pp. 95–100, 2007.

[4] D. Prideaux, C. Roberts, K. Eva et al., "Assessment for selection for the health care professions and specialty training: consensus statement and recommendations from the Ottawa 2010 conference," *Medical Teacher*, vol. 33, no. 3, pp. 215–223, 2011.

[5] H. Coates, "Establishing the criterion validity of the Graduate Medical School Admissions Test (GAMSAT)," *Medical Education*, vol. 42, no. 10, pp. 999–1006, 2008.

[6] D. Edwards, T. Friedman, and J. Pearce, "Same admissions tools, different outcomes: a critical perspective on predictive validity in three undergraduate medical schools," *BMC Medical Education*, vol. 13, article 173, 2013.

[7] A. Mercer and I. B. Puddey, "Admission selection criteria as predictors of outcomes in an undergraduate medical course: a prospective study," *Medical Teacher*, vol. 33, no. 12, pp. 997–1004, 2011.

[8] P. Poole, B. Shulruf, J. Rudland, and T. Wilkinson, "Comparison of UMAT scores and GPA in prediction of performance in medical school: a national study," *Medical Education*, vol. 46, no. 2, pp. 163–171, 2012.

[9] D. Wilkinson, J. Zhang, and M. Parker, "Predictive validity of the undergraduate medicine and health sciences admission test for medical students' academic performance," *Medical Journal of Australia*, vol. 194, no. 7, pp. 341–344, 2011.

[10] I. C. McManus, C. Dewberry, S. Nicholson, and J. S. Dowell, "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 12 UK medical schools," *BMC Medicine*, vol. 11, article 244, 2013.

[11] E. Siu and H. I. Reiter, "Overview: what's worked and what hasn't as a guide towards predictive admissions tool development," *Advances in Health Sciences Education*, vol. 14, no. 5, pp. 759–775, 2009.

[12] K. W. Eva, H. I. Reiter, J. Rosenfeld, and G. R. Norman, "The ability of the multiple mini-interview to predict preclerkship performance in medical school," *Academic Medicine*, vol. 79, no. 10, pp. S40–S42, 2004.

[13] F. Lievens and P. R. Sackett, "Video-based versus written situational judgment tests: a comparison in terms of predictive validity," *Journal of Applied Psychology*, vol. 91, no. 5, pp. 1181–1188, 2006.

[14] C. D. Kreiter and R. D. Axelson, "A perspective on medical school admission research and practice over the last 25 years," *Teaching and Learning in Medicine*, vol. 25, supplement 1, pp. S50–S56, 2013.

[15] K. W. Eva, H. I. Reiter, K. Trinh, P. Wasi, J. Rosenfeld, and G. R. Norman, "Predictive validity of the multiple mini-interview for selecting medical trainees," *Medical Education*, vol. 43, no. 8, pp. 767–775, 2009.

[16] K. W. Eva, H. I. Reiter, J. Rosenfeld, K. Trinh, T. J. Wood, and G. R. Norman, "Association between a medical school admission process using the multiple mini-interview and national licensing examination scores," *The Journal of the American Medical Association*, vol. 308, no. 21, pp. 2233–2240, 2012.

[17] A. Husbands and J. Dowell, "Predictive validity of the Dundee multiple mini-interview," *Medical Education*, vol. 47, no. 7, pp. 717–725, 2013.

[18] H. I. Reiter, K. W. Eva, J. Rosenfeld, and G. R. Norman, "Multiple mini-interviews predict clerkship and licensing examination performance," *Medical Education*, vol. 41, no. 4, pp. 378–384, 2007.

[19] F. Lievens, T. Buyse, and P. R. Sackett, "The operational validity of a video-based situational judgment test for medical college admissions: illustrating the importance of matching predictor and criterion construct domains," *Journal of Applied Psychology*, vol. 90, no. 3, pp. 442–452, 2005.

[20] F. Lievens and P. R. Sackett, "The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance," *Journal of Applied Psychology*, vol. 97, no. 2, pp. 460–468, 2012.

[21] P. R. Sackett and H. Yang, "Correction for range restriction: an expanded typology," *Journal of Applied Psychology*, vol. 85, no. 1, pp. 112–118, 2000.

[22] M. Beller, "Psychometric and social issues in admissions to Israeli universities," *Educational Measurement: Issues and Practice*, vol. 13, no. 2, pp. 12–20, 1994.

[23] N. Gafni, A. Moshinsky, O. Eisenberg, D. Zeigler, and A. Ziv, "Reliability estimates: behavioural stations and questionnaires in medical school admissions," *Medical Education*, vol. 46, no. 3, pp. 277–288, 2012.

[24] A. Ziv, O. Rubin, A. Moshinsky et al., "MOR: a simulation-based assessment centre for evaluating the personal and interpersonal qualities of medical school candidates," *Medical Education*, vol. 42, no. 10, pp. 991–998, 2008.

[25] H. Gulliksen, *Theory of Mental Tests*, John Wiley & Sons, New York, NY, USA, 1950.

[26] C. Oren, T. Kennet-Cohen, E. Turvall, and A. Allalouf, "Demonstrating the validity of three general scores of PET in predicting higher education achievement in Israel," *Psicothema*, vol. 26, no. 1, pp. 117–126, 2014.

[27] T. Kennet-Cohen, E. Turvall, and C. Oren, "Detecting bias in selection for higher education: three different methods," *Assessment in Education: Principles, Policy & Practice*, vol. 21, no. 2, pp. 193–204, 2014.

[28] M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein, *Introduction to Meta-Analysis*, John Wiley & Sons, Chichester, UK, 2009.

[29] P. Bobko and A. Rieck, "Large sample estimators for standard errors of functions of correlation coefficients," *Applied Psychological Measurement*, vol. 4, no. 3, pp. 385–398, 1980.

[30] R. L. Linn, "Pearson selection formulas: implications for studies of predictive bias and estimates of educational effects in selected samples," *Journal of Educational Measurement*, vol. 20, no. 1, pp. 1–15, 1983.

[31] I. C. McManus, K. Woolf, J. Dacre, E. Paice, and C. Dewberry, "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*, vol. 11, no. 1, article 242, 2013.

Autism
Research and Treatment

Economics
Research International

Journal of
Biomedical Education

Nursing
Research and Practice

Child Development
Research

Journal of
Criminology

Journal of
Archaeology

International Journal of
Population Research

Education
Research International

Sleep Disorders

Depression Research
and Treatment

Journal of
Anthropology

Journal of
Addiction

Schizophrenia
Research and Treatment

Current Gerontology
& Geriatrics Research

Geography
Journal

Journal of
Aging Research

Psychiatry
Journal

Urban Studies
Research

Hindawi

Submit your manuscripts at
http://www.hindawi.com