

Overview: what's worked and what hasn't as a guide towards predictive admissions tool development

Eric Siu · Harold I. Reiter

Received: 19 September 2008 / Accepted: 20 March 2009 / Published online: 2 April 2009
© Springer Science+Business Media B.V. 2009

Abstract Admissions committees and researchers around the globe have used diligence and imagination to develop and implement various screening measures with the ultimate goal of predicting future clinical and professional performance. What works for predicting future job performance in the human resources world and in most of the academic world may not, however, work for the highly competitive world of medical school applicants. For the job of differentiating within the highly range-restricted pool of medical school aspirants, only the most reliable assessment tools need apply. The tools that have generally shown predictive validity in future performance include academic scores like grade point average, aptitude tests like the Medical College Admissions Test, and non-cognitive testing like the multiple mini-interview. The list of assessment tools that have not robustly met that mark is longer, including personal interview, personal statement, letters of reference, personality testing, emotional intelligence and (so far) situational judgment tests. When seen purely from the standpoint of predictive validity, the trends over time towards success or failure of these measures provide insight into future tool development.

Keywords Predictive validity · Admissions · Overview · Grade point average · Medical College Admissions Test · Multiple mini-interview

Overview

Reader be warned—the following is not a formal literature review. That need has already been elegantly addressed for non-cognitive measures (Albanese et al. 2003), for grade point average (GPA) (Kreiter and Kreiter 2007) and for subsections of the Medical College Admissions Test (MCAT) (Donnon et al. 2007). This paper is meant as an overview,

E. Siu
McMaster University, 1200 Main Street West, MDCL 3112, Hamilton, ON L8N 3Z5, Canada

H. I. Reiter (✉)
Department of Oncology, McMaster University, 1200 Main Street West, MDCL 3112, Hamilton,
ON L8N 3Z5, Canada
e-mail: reiterh@mcmaster.ca

looking for trends in the predictive validity data provided by admissions tools in the hope that this might provide guidance for future assessment tools development.

In pursuit of cogency, some rules of conduct were set for this overview. Measurement tools were considered to be of helpful predictive validity if they demonstrated all of the following four characteristics:

1. The positive predictive validity correlation must be statistically significant.
2. The positive predictive validity correlation must be practically relevant. A very highly powered study might demonstrate the existence of a correlation with great confidence, and with equally great confidence confirm that the extent of the correlation is vanishingly small.
3. The positive predictive validity correlation must be consistent across multiple studies. It is acknowledged a priori that seeking positive correlations with unreliable outcome variables may limit achievement of this characteristic.
4. The positive predictive validity correlation must have a value added, an incremental validity above and beyond other predictors.

When examining strength of data, three levels, from high to low, are considered: (a) meta-analysis, (b) large studies (>500 subjects) and (c) small studies. Lower level data is considered only when higher level data is unavailable.

What's worked...

This list is limited to GPA, the MCAT and the multiple mini-interview (MMI).

Grade point average

Grade Point Average has consistently shown statistically significant, practically relevant, positive predictive correlations with future performance, confirmed on meta-analysis (Kreiter and Kreiter 2007). It has some degree of incremental validity above the MCAT (Julian 2005). Correlations are particularly strong (0.40), trending downwards with increasing time from medical school admission. Awareness of this downward trend is longstanding (Gough 1978). Confounding factors for this decrease may include changing cognitive content and/or a shift of content from cognitive towards non-cognitive emphasis in outcome variables administered later in training.

No clear conclusions can be drawn that differentiate between overall GPA and science GPA, as the latter predominates as the standard in reported studies, and no head-to-head comparison is readily available (Koenig et al. 1998; Huff et al. 1999). Science GPA continues as the overwhelmingly preferred GPA indicator in American schools and some Canadian schools without clear data support for that phenomenon, and with only occasional challenge (Barr et al. 2008). Several studies suggest that students from non-science backgrounds initially experience higher stress but ultimately perform equally well relative to their science background counterparts (Dickman et al. 1980; Yens and Stimmel 1982; Woodward and McAuley 1983; Neame et al. 1992; Koenig 1992; Huff and Fang 1999). These studies, however, are plagued by the use of unreliable outcome variables. A lack of demonstrated difference may be because no such difference exists; alternatively, a difference exists, but cannot demonstrate correlation with random number generators.

Medical college admissions test

The MCAT has consistently been shown to be statistically significant, practically relevant, with a positive predictive correlation with future performances. This conclusion has been confirmed on meta-analysis (Donnon et al. 2007). The MCAT has a strong degree of incremental validity, above and beyond GPA (Julian 2005). Correlations are particularly strong, trending downwards with increasing time from medical school admission, and with increasing shift of focus from predominantly cognitive outcome variables earlier in training to predominantly non-cognitive and clinical outcome variables later in training. These trends, however, vary considerably for different MCAT sections.

The MCAT itself is apportioned into four sections. The Physical Sciences section is intended to assess problem solving ability in general chemistry and physics; the Biological Sciences section is intended to do the same for organic chemistry and biology. The Writing Sample section requires the composition of two essays, intended to measure candidates' ability to develop a central idea, synthesize concepts, and present those ideas cohesively, logically, and with correct use of grammar and syntax. The Verbal Reasoning section consists of approximately seven passages, each followed by 5–7 questions, whose correct response requires the candidate to understand, evaluate and apply the information and arguments provided.

Physical Sciences (MCAT-PS) demonstrates a moderately strong correlation that drifts downwards with increasing time from medical school admission.

Biological Sciences (MCAT-BS) demonstrates a strong correlation which trends downwards with increasing time from medical school admission, though in an even less precipitous fashion than MCAT-PS.

Writing Sample (MCAT-WS) would be categorized more appropriately in the section below entitled "and What Hasn't (Worked)". MCAT-WS has failed to demonstrate consistently positive results for predictive validity.

Verbal Reasoning (MCAT-VR) demonstrates moderately strong correlations. These correlations are not only sustained, but strengthened, with increasing time from medical school admission. The relative immunity to time of MCAT-VR compared to MCAT-PS and MCAT-BS (Violato and Donnon 2005) bears some scrutiny. It may be due to a combination of factors. One factor may be that MCAT-VR is less context-bound, so correlations with future performance remain unaffected as the context changes. Another may be that MCAT-VR straddles the somewhat artificial divide between cognitive and non-cognitive domains and therefore remains relevant, even as assessment measures in clerkship and on national licensing examinations shift from cognitive towards non-cognitive domains.

Multiple mini-interview

The MMI has consistently shown statistically significant, practically relevant, positive predictive correlations with future performance. This conclusion has been confirmed on small studies only (Eva et al. 2004, 2009; Reiter et al. 2007). It has a strong degree of incremental validity above GPA and MCAT. This is unsurprising, given the zero correlation of MMI with GPA and only small correlation of MMI with MCAT-VR. Correlations are sustained and trend upwards with increasing time from medical school admission. This may be due to a shift from cognitive towards non-cognitive domains in later outcome assessments during clerkship and national licensure examination. Conclusions are guarded,

pending assessment of a sufficiently large cohort of MMI-tested individuals reaching USMLE Steps I, II and III.

...And what hasn't...

This list includes personal interviews (PI), personal statements, letters of reference (LOR), personality testing (PT), measures of emotional quotient/emotional intelligence (EQ/EI), and written and video-based situational judgment tests (w-SJT and v-SJT).

Personal interview

Personal interviews should work. They certainly work in the Human Resources (HR) setting. The HR methodological gold standard includes a pre-interview job analysis, behavioural descriptor interview (BDI) questions (e.g. describe an occasion you felt challenged in the workplace, and how you dealt with that challenge), and situational interview (SI) questions (e.g. faced with the following challenge, how would you deal with it). This approach yields predictive validity correlations of 0.20–0.30 with subsequent job performance (Wiesner and Cronshaw 1988; Taylor and Small 2002). Yet similar results have not been found for medical school admissions interviews (Albanese et al. 2003; Salvatori 2001).

Personal interview for medical school admission has demonstrated positive predictive validity in three studies, one published after the Albanese review. However, the personal interview still fails to meet the criteria described at this paper's outset. Firstly, the correlations in two studies (Powis et al. 1988; Kreiter et al. 2004) were not found for the entire cohort under study, but only found for applicants scoring extremely highly and extremely poorly in both predictor and outcome variables. Further, the Kreiter study employed a high level of standardized interview methodology rarely achieved by other medical schools. Short of marked enhancement of other medical schools' interview format, the generalizability of these results may therefore be questioned. Most intriguing, however, is that two studies (Meredith et al. 1982; Powis et al. 1988) employed a variation in the interview format that likely contributed to serendipitous results. Specifically, Meredith et al. used four independent interviews and Powis et al. used two independent interviews rather than one, for each candidate. As with MMI, this would tend to combat the negative impact of context specificity and halo effect on overall test reliability. The Meredith study has the previously unrecognized distinction of being the first published study to demonstrate some degree of positive predictive validity for a multiple interview, a full generation prior to Eva et al. and the MMI.

Why the difference in results between HR and medical school interviews? The latter tend not to use BDI/SI questions, a factor which would drop predictive validity to the 0.10–0.20 range, according to the HR experience (Wiesner and Cronshaw 1988; Huffcutt and Arthur 1994; McDaniel et al. 1994). But even weak correlations have not been consistently achieved by medical schools. The difference likely lies in differing degrees of applicant pool homogeneity for HR and medical schools. The smaller the difference between individual applicants, the greater the challenge in differentiating between them, and the less likely there will be wide differences in the hired/admitted applicants' subsequent job performance.

For better or for worse, the practice of medicine has a higher profile than most job positions. Those who apply in a casting call for leads in a Broadway play are going to

appear very much different from those who apply for an off-Broadway play. The resumes of the stars on Broadway will cluster at the upper end of the scale, and the theatre-goers can rightly expect a uniformly expert performance. The resumes of the off-Broadway group are across the board and the viewing audience knows in advance that the performance will vary greatly based upon the chosen performers. In a similar vein, it is far more challenging to differentiate between medical school applicants and subsequently demonstrate positive predictive validity, as compared to the usual HR cattle calls.

Letters of reference

Letters of references (LOR) have been an integral part of the medical school admissions process and remains as one of the most common criteria in the candidate screening process (DeLisa et al. 1994; Berstein et al. 2002). There is, however, little evidence to support their effectiveness and their continued usage in the medical school admission process (Salvatori 2001). Many of the concerns that arise from the use of LOR as a selection tool can be attributed to its poor predictive validity (Kirchner and Holm 1997; Standridge et al. 1997) and poor reliability (Ross and Leichner 1984). In terms of inter-rater reliability, Dirschl and Adams (2000) evaluated LOR for 58 residency applicants and found that inter-rater reliability was slight (0.17–0.28). Concerns raised regarding the use of LOR as part of the admission process also focus on the lack of information in these documents, a perception of ambiguity with terminology and rater bias as a result of open file reviews.

Although it is widely believed that traditional LOR offer a great deal of information about an applicant's non-cognitive abilities, little has been found to support this contention. In a study conducted by O'Halloran et al. (1993), two experienced reviewers could not reliably extract information on non-cognitive qualities from candidate letters of reference. Ferguson et al. (2003) has voiced similar concerns stating that the amount of information contained in a teacher's reference does not reliably predict the performance of a student at medical school.

In his review of Dean's reference letters for pediatric residency programs, Ozuah (2002) found that a substantial proportion of Deans' letters from US medical schools failed to include comparative performance information with a key that allowed for accurate interpretation. Such subjective letters can be one of the many reasons that contribute to the poor inter-rater reliability highlighted above.

Interestingly, both studies by Brothers and Wetherholt (2007) and Peskun et al. (2007) found letters of reference to be predictive of later performance. Brothers and Wetherholt (2007) found a correlation with subsequent clinical performance ratings in residency training. However, the authors acknowledge an important limitation in this situation in that faculty raters also received information about academic performance and USMLE scores that accompanied the reference letters. As such, this open file review process likely biased their overall scoring of the applicant's LOR. The same may or may not be true for Peskun et al. (2007), who found that non-cognitive assessments (LOR and personal statements culled from medical school admissions files) provided additional value to standard academic criteria in predicting ranking by two residency programs.

Personal statements

Much like the assessment of LOR, there has also been little support found for the predictive validity of personal statements (McManus and Richards 1986). Ferguson et al.

(2000) evaluated the personal statements of 176 medical students and found that neither the information categories nor the amount of information in the statements were predictive of future preclinical performance.

Personal statements are also subject to rater biases, which can have notable effects on reliability. In a study looking at the autobiographical sketch submission of applicants to the Michael G. Degroote School of Medicine at McMaster University, Kulatunga-Moruzi and Norman (2002) reported an intermediate inter-rater reliability coefficient of 0.45. Later research (Dore et al. 2006; Hanson et al. 2007) suggested that even that modest claim might be unreasonably optimistic.

Personal statements add relatively little value to an applicant's profile, due to the common occurrence of input from others in an applicant's statement, as well as a subjective and difficult comparison of personal statements across applicants. In surveys of first year medical students conducted over 3 years, Albanese et al. (2003) found that 41–44% of students reported their personal statement involved input from others with 15–51% reporting input in content development and 2–6% receiving input from professional services. Similarly, in an assessment of off-site pre-interview autobiographical submissions compared with onsite submissions, Hanson et al. (2007) suggested a relatively low level of frequency with which candidates independently answered pre-interview autobiographical questions, and reported on its deleterious impact on test reliability. This is hardly surprising. It is in the best interest of all applicants to present themselves in a fashion that clusters them at the saintly end of the spectrum. The higher the stakes, the tighter the cluster.

Albanese et al. (2003) has additionally pointed out that because of the personal statement's free-form nature, "any given personal statement will highlight a set of personal characteristics potentially different from the set highlighted in another applicant's personal statement". Comparing non-standardized information of applicants' personal characteristic makes valid comparisons extremely challenging.

Although research has indicated limited predictive value of the personal statement as a selection tool, the focus of this tool in medical school admissions literature has remained sparse. Interestingly, there have been conflicting reports on the reliability and validity of personal statements in the admissions processes in other health care disciplines. Kirchner and Holm (1997) found a positive correlation between the autobiographical essay and Occupational Therapy GPA. The essay was also found to have added incremental validity to their model used to predict therapy outcomes. However, in a follow up study, Kirchner et al. (2000) could not replicate the same positive correlation and suggested that the previous results may have been anomalous. Similarly, Brown et al. (1991) reported good inter-team reliability (0.71–0.80) when assessing applications to the basic stream of the baccalaureate nursing program at McMaster University. When evaluating letters of applicants applying to the post RN stream, however, they found a lower coefficient of 0.43. The authors attributed some of this variance to rater bias.

Although there is conflicting data regarding the reliability and validity of personal statements in other health care disciplines, the influence of bias and random error cannot be ruled out. Additionally, the higher stakes of medical school compared to other health care schools will tend to cluster the former to haloed extremes. If anything, then, weak results for other health disciplines augurs even more poorly for medical school admissions. Not surprisingly, without meeting the outlined criteria of being replicable across studies, personal statements cannot be considered as an assessment tool that "works".

Personality testing

A wealth of large scale studies, literature reviews and meta-analyses can be found on personality testing in the human resource (HR) literature (Barrick and Mount 1991). That literature provides a source of encouragement for the application of PT to medical school admissions, albeit with a rather large caveat. As discussed above, the population of medical school aspirants tends to be far more homogenous than those reported in HR studies. A microscope strong enough to differentiate between HR job applicants may not be anywhere near powerful enough to do the same for medical school applicants. Even worse, people are generally unable to reliably self-report. Self-assessment tends to be profoundly inaccurate, so it is likely expecting too much to have self-reporting tools like personality tests avoid myopic, albeit consistent, viewpoints. The automobile driver who knows that he/she is a good driver will consistently report themselves as such, regardless of true driving ability. Because of this, high internal consistency or Cronbach's alpha using personality tests should not enter the discussion—there is no great merit to being consistent in one's responses when those responses are consistently inaccurate. This conclusion in no way negates the successful use of personality testing under the auspices of forensic psychologists, who deal with a pool of individuals liberally sprinkled with psychopaths and sociopaths. Medical school applicant pools are homogeneously populated by high academic performers, with psychopaths sufficiently (and thankfully) rarely represented, so as to make the personality test unreliable in that setting.

Part of the challenge in interpreting results of early personality testing was generated by the plethora of different personality testing constructs. Using different constructs, with different measuring tools found within each construct, and translating between study results and generalizing from one set of results was an exercise in futility. Over time, one construct—the Big Five Factor construct—and one measuring tool within that construct—NEO (Neuroticism-Extroversion-Openness) Five Factor Inventory—have garnered more widespread use. The NEO Five Factor Inventory is composed of Conscientiousness, Openness to Experience, Neuroticism, Extroversion, and Agreeableness. Of these, only Conscientiousness has consistently demonstrated predictive validity. Claims of predictive validity of the other factors are plagued by failure to make Bonferroni corrections when many correlations of multiple endpoints are sought. If a study finds that a proportion close to 5% of the correlations sought are statistically significant at $P < 0.05$, then the “positive result” is quite likely due to random chance. Put another way, let's say Mr. Q boasts of his betting system on horses, and returns from the track having won his bet on a horse running at 20:1 odds. Would you bet money on his system? Well, did he win that on a single bet, or did he win one 20:1 bet while losing 19 others at the same odds? Further, if he tells you that he won that bet because the winning horse had the longest legs and then tells you a week later that he has won another 20:1 bet because the winning horse had the most experienced jockey, you might be wise to keep your wallet closed. Explaining the reasons for the winning bet only after the fact, and rarely in consonance with other post hoc explanations of winning bets, is another warning sign. The HR literature is replete with such situations of positive correlations in one study explained sagaciously, without similarly robust findings in other studies. For these reasons, only the NEO factor of Conscientiousness demonstrates sufficiently consistent findings in order to have it described, along with aptitude tests, as worthwhile predictive measures to consider (Behling 1998).

The NCQ (non-cognitive questionnaire) is another example of a personality testing construct which has gained a level of acceptance, best detailed in *Beyond the Big Test*:

Non-cognitive Assessment in Higher Education, by William Sedlacek (Sedlacek 2004). Attempting to interpret the conclusions drawn is a challenge. Trying to extrapolate a concept that is pertinent to medical school admissions from these conclusions is an even more daunting task. Of the extensive list of 403 references cited in the book, 15 address predictive validity results in peer-reviewed journals (Bandalos and Sedlacek 1989; Boyer and Sedlacek 1988; Fuertes and Sedlacek 1994, 1995; O'Callaghan and Bryant 1990; Sedlacek 1991, 1996, 1999, 2003; Sedlacek and Adams-Gaston 1992; Ting 1997; Tracey and Sedlacek 1985, 1987, 1988, 1989). The vast majority of these 15 publications do not deal with overall applicant populations, but rather subgroups, particularly racially defined subgroups. Are the results generalizable, particularly when applied to the medical school applicant population, who are likely to be far more homogenous than the populations examined using NCQ? Do the number of statistically significant positive predictive correlations exceed that which one would expect by chance alone? Further to this issue, there is no explicit use of Bonferroni corrections. Like other personality tests, those positive correlations that are found are explained post hoc, and are not necessarily consistent between studies. Finally, when statistically significant associations between NCQ scores and academic scores are found, it is not always inherently clear which represents the predictor variable and which represents the outcome variable. As acknowledged by Ting (1997),

“Psychosocial variables including successful leadership experiences, preference for long-range goals, acquired knowledge in a field and a strong support person were significantly related to GPA in the 1st and 2nd semesters. Thus, students who have higher scores on these psychosocial variables also tend to have higher GPAs, or vice versa.”

If the score for “availability of a strong support person” correlates with higher GPA, is that because the former led to the latter, or because a strong support person is more likely to be attracted to those more likely to succeed academically? Ultimately, the limitations in interpreting and extrapolating NCQ results do not necessarily mean that the construct and tests proposed are without merit; only that it is difficult to draw conclusions based upon the extensive information provided.

Emotional quotient/emotional intelligence

Unlike personality testing, testing of EQ/EI has not developed to the point that a single test format has gained greater favour over most others. Worse, it has not developed to the point that a single construct has gained greater favour over most others. With the incongruence of different constructs and different test formats essentially speaking different languages, the interpretation of results between studies is all but impossible. Furthermore, all the same limitations expected when interpreting personality testing studies for their applicability to medical school admissions are also present in emotional intelligence studies: attempted extrapolations from results with more heterogeneous populations, misplaced trust in ability to self-assess, lack of Bonferroni corrections and lack of robust results between studies, even when similar constructs and test formats can be found. The promise and the challenge of EQ/EI has been recently addressed by Lewis et al. (2005), and will not be further addressed here, beyond accepting that the existing predictive validity data does not provide a compelling argument for its present use.

Written and video based situational judgment tests

Recently, situational judgment tests (SJT) administered during the medical school admissions process have emerged as relatively strong predictors of future academic performance (Lievens et al. 2005; Lievens and Sackett 2006; Oswald et al. 2004). In SJTs, applicants are presented with either written or video based depictions of hypothetical scenarios and are asked to identify an appropriate response from a list of alternatives (Lievens et al. 2005). The intent was to develop an admissions test which might predict for future non-cognitive performance.

In their evaluations of SJT, Lievens and Sackett (2006) examined the validity of written based versus video based SJT in a traditional student admission tests for candidates writing the Medical and Dental Studies Admission Exam in Belgium. They compared 1,159 students who completed the video SJT against 1,750 who completed the same SJT but in a written format. It was found that the written SJT offered significantly less predictive and incremental validity for predicting interpersonally oriented criteria than did the video SJT. Interestingly, the written SJT was more predictive of cognitive aspects of future performance as measured by GPA (Lievens and Sackett 2006). As previously discussed in this paper, GPA and MCAT scores have already been proven to be strong cognitively oriented predictors. Thus, since the intention of the SJT was to provide alternative measures to assess non-cognitive qualities, the written SJT has little added value.

Contrastingly, video SJT have shown very promising results in its overall ability to predict future performance. Video SJT, as mentioned earlier, had significantly higher predictive validity and incremental validity for predicting interpersonally oriented criteria than did the written SJT (Lievens and Sackett 2006). Specifically, in their sampling of 1,159 candidates Lievens et al. (2005), found that the video based SJT proved to be a significant predictor in academic courses that partially determined GPA based on interpersonally oriented courses with a positive correlation of 0.21 between the SJT validity coefficient and the interpersonal course ratings. In contrast to the written SJT, the video based version was not predictive for cognitive domains. Further, the video SJT validity increased through the academic years when following students through their first 4 years of medical training (Lievens et al. 2005).

One major limitation that Lievens et al. (2005) address is the differences between admission for North American and European medical schools. It is worth noting that not only is the admission process different, whereby medical school admission is controlled by the state in Belgium, but the population of medical school aspirants also tends to be far more homogenous in North America than in Belgium. Applicants to the majority of North American medical schools are required to have completed several years of post-secondary studies before applying to medical school. Most prospective applicants with lower academic achievement have been siphoned off, leaving a more homogenous applicant pool. The majority of applicants thus have very competitive profiles making selection criteria that much more difficult amongst such a homogeneous applicant pool. This may be very different from the more heterogeneous candidates in Belgium applying to medicine right after secondary studies. It is as yet unclear whether applying the video SJT methodology to the more homogenous (more highly clustered) North American applicant pool maintains a sufficient level of predictive validity to warrant its ongoing use. Since video SJT does not correlate with GPA, it remains possible, albeit unconfirmed, that this homogenizing or clustering effect is minimal for non-cognitive domains.

Because there are not as of yet multiple studies demonstrating predictive validity, video SJT, for the moment, remains in the "and what has not" portion of this paper. The

likelihood of future widespread implementation may also be limited by the same cost and technological concerns that led to cessation of video SJT use in Belgium in 2002. Nevertheless, the promising results from Belgium may yet herald a future move of video SJT into the “what’s worked” portion of this paper.

As a guide...

As stated earlier, the overview is looking for trends in the predictive validity data provided by admissions tools in the hope that this might provide guidance for future assessment tool development. Yes, Grade Point Average provides statistically significant positive predictive validity, but how do those correlations change both over time and over the shifting emphasis of endpoints from purely cognitive to clinical? We may be happy that those admitted are more likely to perform well in basic medical science courses in the first 2 years of medical school, but can that endpoint hold a candle to performance as clinical clerks? Does clerkship performance carry the same cachet as scores on certification examinations? Not when higher certification examination scores predict the appropriate use of radiological investigations and correct prescribing techniques (Tamblyn et al. 1998, 2002), lower mortality following cardiac events (Norcini et al. 2002), peer ratings of clinical competence (Ramsey et al. 1989) and fewer complaints to medical boards (Tamblyn et al. 2007). The shift in emphasis can be illustrated in tabular and graphic form, moving from earlier to later training, from mainly cognitive to mainly clinical assessment and ultimately to assessment of professionalism. With that in mind, the endpoints available in the literature are presented in the following series, in order:

1. Grades in the first 2 years of medical school (1st 2 years).
2. Results of early (less clinical) portions of national licensure examinations (US Medical Licensing Examination Steps 1 and 2, Medical Council of Canada Qualifying Examination Part I; USMLE 1, 2, MCC I).
3. Results of early portions of national licensure examinations explicitly dealing with clinical issues (Medical Council of Canada Qualifying Examination Part I—Clinical Decision-Making; MCC I CDM).
4. Clinical Clerkship scores.
5. Results of later portions of national licensure examinations dealing to a greater extent with clinical issues (US Medical Licensing Examination Step 3, Medical Council of Canada Qualifying Examination Part II; USMLE 3, MCC II).
6. Results of later portions of national licensure examinations explicitly dealing with clinical issues (Medical Council of Canada Qualifying Examination Part II—Clinical Decision-Making; MCC II CDM).
7. Results of national licensure examinations explicitly dealing with issues of professionalism (Medical Council of Canada—Considerations of Legal, Ethical and Organizational; MCC-CLEO).

For those tools that have not reproducibly demonstrated positive predictive validity, there is no trend to show, and they are excluded from the table and graphs. The tools that are included in tabular (see Table 1) and linear trend graphic form include GPA (see Fig. 1), the three predictive sections of MCAT (see Figs. 2, 3, 4), and MMI (see Fig. 5). The numbers used to fill in the table arise from (a) meta-analyses (designated in bold), (b) large studies (>500 subjects; designated in regular font), and (c) small studies (designated in italics).

Table 1 Predictive tools and their correlation with future assessments

	1st 2 years	USMLE 1, 2, MCC I	MCC I CDM	Clerk	USMLE 3, MCC II	MCC II CDM	MCC CLEO
GPA	0.4	0.35	0.25	0.26	0.29	0.19	0.09
WS	-0.13	0.07		0.07		-0.06	
PS	0.23	0.36		0.06		0.02	
BS	0.32	0.39		0.12	0.11	0.03	
VR	0.19	0.27		0.14	0.27	0.24	0.43
MMI	0.18	0.17	0.18	0.57	0.35		0.37

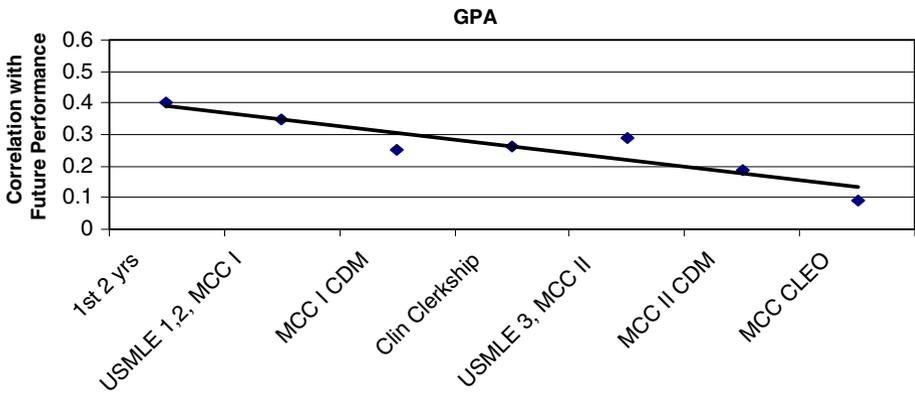


Fig. 1 Correlation of GPA with future assessments

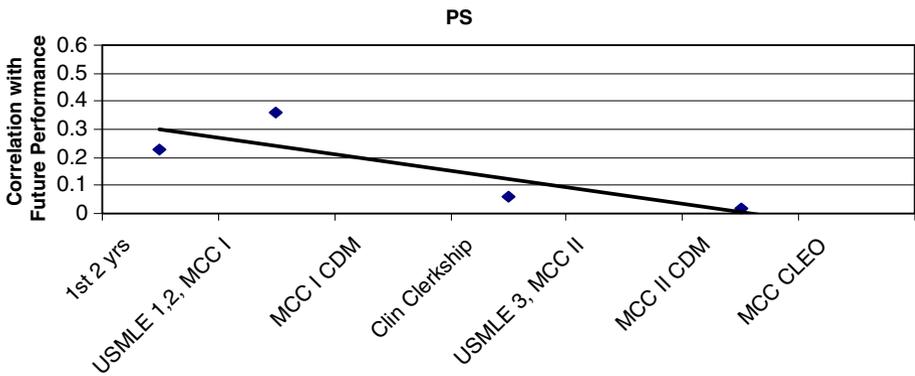


Fig. 2 Correlation of MCAT-PS with future assessments

Towards predictive admissions tool development

Grade Point Average predicts for Grade Point Average. Correlation between one course score and another, whether in or out of medical school, is moderate; correlation between

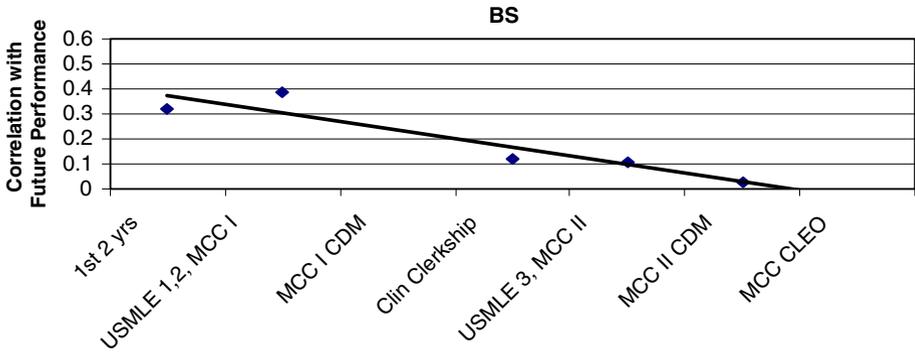


Fig. 3 Correlation of MCAT-BS with future assessments

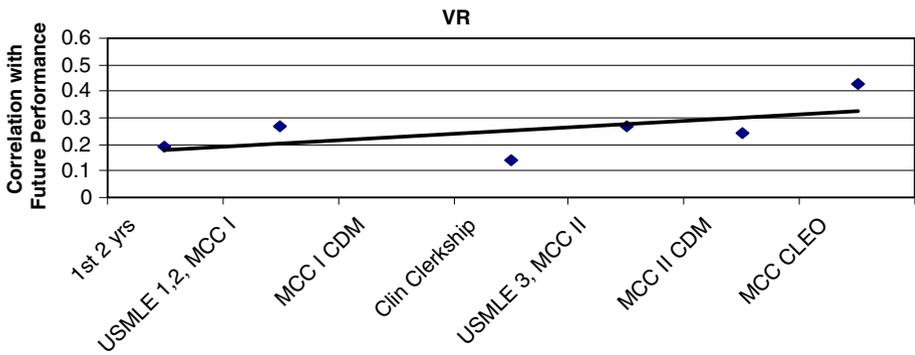


Fig. 4 Correlation of MCAT-VR with future assessments

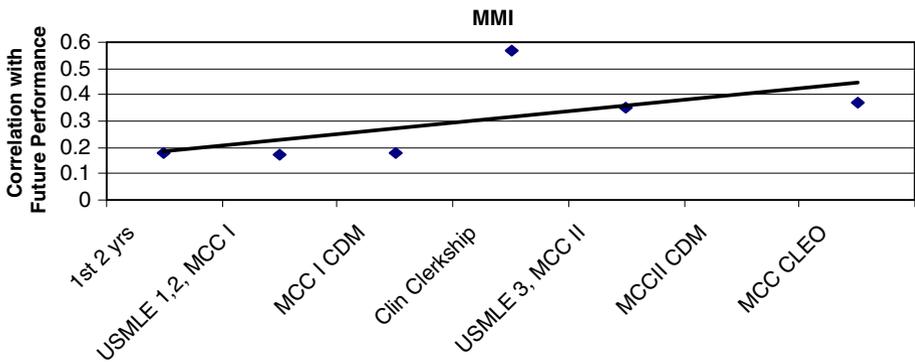


Fig. 5 Correlation of MMI with future assessments

years of courses is extremely high (Trail et al. 2006). But the ability to excel academically carries less and less gravitas as the domain assessed shifts from the more purely cognitive to the more clinical and ultimately more professional. If the goal of medical schools is to churn out medical science cognitive experts, then GPA is the way to go. The real world,

however, places a higher premium on the superb clinician and professional—at least, so it would seem. But in that real world, there are not a lot of physicians with weak cognitive skills. The majority of complaints to State medical boards may be due to issues of professionalism, but that is only because the vast majority of medical aspirants of lower intellectual caliber have been weeded out by GPA and by the Biological Science and Physical Science sections of the MCAT whose predictive validity trends mirror those of GPA (see Figs. 1, 2, 3). Without these screening measures, a much higher proportion of complaints would be due to cognitive, rather than professional, ineptness.

Society has been saved from that fate by the insistence of strong cognitive ability, the potential to attain expertise in medical science, and owes much to the Flexner Report of 1910 in that regard. GPA and MCAT have proven sufficiently successful that further progress on cognitive assessment would provide only limited returns. But scaling one mountain peak opens vistas on the challenges beyond. In a recent sampling of three state medical boards, unprofessional behaviour represented 92% of the known types of violations (Papadakis et al. 2005). This figure is proportional to expectations culled from surveys of sued physicians, non-sued physicians, and suing patients (Shapiro et al. 1989). The approach of the centennial of Flexner's signature report provides increasing perspective on the next domain to be challenged and conquered. Without Flexner, we would still be mired in complaints of cognitive origin. In response to Flexner, the cognitive mountain peak has been scaled and with it, the vantage point to see the next mountain peak to be assaulted. Where is the Flexner Report of 2010, to define non-cognitive skills and in particular professionalism as the next mountains to scale?

How is that assault to be managed? Success in conquering the first, cognitive peak and early success in non-cognitive evaluation provide perspective (TAU) on potential future success—Trust no-one, Avoid self-reporting, Use repeated measures:

1. Trust no-one. Do not trust the applicants. When stakes are high, the likelihood of cheating increases; the stakes for medical school admission are very high. Do not trust the referees; they were not chosen by the applicants for their ability to remain neutral and objective. Do not trust the file reviewers; compared with the formulaic approach, individual biases always get in the way, from parole board decisions (Grove and Meehl 1996) to medical school admissions (Schofield and Garrard 1975). Do not trust the interviewers, or at the very least, dilute their biases into submission.
2. Avoid self-reporting. Applicants' inability to cheat self-reporting instruments intentionally is small comfort when they can so easily fill them out inaccurately when telling the truth. It's not just that people are generally poor at self-assessment, but also that the worst performers tend to be the worst self-assessors (Eva and Regehr 2007). To expect the bad apples to weed themselves out on the basis of self-reporting goes beyond Pollyanna into self-delusion. All observations about the candidate should be made from a more remote position than the applicants themselves.
3. Use repeated measures. Medical schools would not admit students based upon one course's GPA, nor would they trust an aptitude test with an insufficient numbers of questions. One interview makes no sense. From Meredith through Powis to Eva, the interview formats that suggest reliability and validity are those that use multiple samples, the more the better. Single measures, however obtained, are indefensible whether viewed on the basis of psychometric principles of context specificity and halo effect, or viewed on the basis of empiric data.

In truth, this prescription for test development has already been applied in limited fashion. In Belgium, video-based situational judgment testing has provided initial success;

in Canada, MMIs have provided the same. As similar endeavours spread globally, it is only a matter of time (and concerted effort) before non-cognitive assessments are taken for granted in the same way that GPA and MCAT are today. That hope leads inexorably to the next obvious question. The conquest of cognitive medical expertise provided a vantage point of the next peak to conquer. As the next peak of non-cognitive skills is scaled, what new peak(s) might become visible? And will it take a century to identify those next challenges and manage the assault?

References

- Albanese, M. A., Snow, M. H., Skochelak, S. E., Huggett, K. N., & Farrell, P. M. (2003). Assessing personal qualities in medical school admissions. *Academic Medicine*, *78*(3), 313–321. doi:[10.1097/00001888-200303000-00016](https://doi.org/10.1097/00001888-200303000-00016).
- Bandalos, D., & Sedlacek, W. (1989). Predicting success of pharmacy students using traditional and non-traditional measures by race. *American Journal of Pharmaceutical Education*, *53*, 145–148.
- Barr, D. A., Gonzalez, M. E., & Wanat, S. F. (2008). The leaky pipeline: factors associated with early decline in interest in premedical studies among underrepresented minority undergraduate students. *Academic Medicine*, *83*(5), 503–511.
- Barrick, M. R., & Mount, M. K. (1991). The Big 5 personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1–26. doi:[10.1111/j.1744-6570.1991.tb00688.x](https://doi.org/10.1111/j.1744-6570.1991.tb00688.x).
- Behling, C. (1998). Employee selection: Will intelligence and conscientiousness do the job? *Academy of Management Executive*, *12*(1), 77–86.
- Berstein, A. D., Jazrawi, L. M., Elbeshbeshy, B., Della Valle, C. J., & Zukerman, J. D. (2002). An analysis of orthopaedic residency selection criteria. *Bulletin (Hospital for Joint Diseases (New York, NY))*, *61*(1–2), 49–57.
- Boyer, S. P., & Sedlacek, W. E. (1988). Noncognitive predictors of academic success for international students: A longitudinal study. *Journal of College Student Development*, *29*, 218–222.
- Brothers, T. E., & Wetherholt, S. (2007). Importance of the faculty interview during the resident application process. *Journal of Surgical Education*, *64*(6), 378–385. doi:[10.1016/j.jsurg.2007.05.003](https://doi.org/10.1016/j.jsurg.2007.05.003).
- Brown, B., Carpio, B., & Roberts, J. (1991). The use of an autobiographical letter in the nursing admissions process: Initial reliability and validity. *Canadian Journal of Nursing Research*, *23*(2), 9–20.
- DeLisa, J. A., Jain, S. S., & Campagnolo, D. I. (1994). Factors used by physical medicine and rehabilitation residency training directors to select their residents. *American Journal of Physical Medicine and Rehabilitation*, *73*(3), 152–156. doi:[10.1097/00002060-199406000-00002](https://doi.org/10.1097/00002060-199406000-00002).
- Dickman, R. L., Sarnacki, R. E., Schimpfhauser, F. T., & Katz, L. A. (1980). Medical students from natural science and nonscience undergraduate backgrounds. Similar academic performance and residency selection. *Journal of the American Medical Association*, *243*(24), 2506–2509. doi:[10.1001/jama.243.24.2506](https://doi.org/10.1001/jama.243.24.2506).
- Dirschl, D. R., & Adams, G. L. (2000). Reliability in evaluating letters of recommendation. *Academic Medicine*, *75*(10), 1029. doi:[10.1097/00001888-200010000-00022](https://doi.org/10.1097/00001888-200010000-00022).
- Donnon, T., Paolucci, E. O., & 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. *Academic Medicine*, *82*(1), 100–106. doi:[10.1097/01.ACM.0000249878.25186.b7](https://doi.org/10.1097/01.ACM.0000249878.25186.b7).
- Dore, K. L., Hanson, M., Reiter, H. I., Blanchard, M., Deeth, K., & Eva, K. W. (2006). Medical school admissions: Enhancing the reliability and validity of an autobiographical screening tool. *Academic Medicine*, *81*(10, Suppl), S70–S73. doi:[10.1097/01.ACM.0000236537.42336.f0](https://doi.org/10.1097/01.ACM.0000236537.42336.f0).
- Eva, K. W., & Regehr, G. (2007). Knowing when to look it up: A new conception of self-assessment ability. *Academic Medicine*, *82*(10, Suppl), S81–S84. doi:[10.1097/ACM.0b013e31813e6755](https://doi.org/10.1097/ACM.0b013e31813e6755).
- Eva, K. W., Reiter, H. I., Rosenfeld, J., & Norman, G. R. (2004). The ability of the multiple mini-interview to predict preclerkship performance in medical school. *Academic Medicine*, *79*(10, Suppl), S40–S42. doi:[10.1097/00001888-200410001-00012](https://doi.org/10.1097/00001888-200410001-00012).
- Eva, K. W., Reiter, H. I., Trinh, K., Wasi, P., Rosenfeld, J., & Norman, G. R. (2009). Predictive validity of the multiple mini-interview for undergraduate and postgraduate admissions decisions. *Medical Education* (accepted for publication).

- Ferguson, E., Sanders, A., O'Hehir, F., & James, D. (2000). Predictive validity of personal statement and the role of the five factor model of personality in relation to medical training. *Journal of Occupational and Organizational Psychology*, 73, 321–344. doi:[10.1348/096317900167056](https://doi.org/10.1348/096317900167056).
- Ferguson, E., James, D., O'Hehir, F., & Sanders, A. (2003). Pilot study of the roles of personality, references, and personal statements in relation to performance over the five years of a medical degree. *British Medical Journal*, 326, 429–432. doi:[10.1136/bmj.326.7386.429](https://doi.org/10.1136/bmj.326.7386.429).
- Flexner, A. (1910). *Medical education in the United States and Canada. Bulletin number four*. New York, NY: Carnegie Foundation for the Advancement of Teaching.
- Fuertes, J. N., & Sedlacek, W. E. (1994). Predicting the academic success of Hispanic university students using SAT scores. *College Student Journal*, 28, 350–352.
- Fuertes, J. N., & Sedlacek, W. E. (1995). Using noncognitive variables to predict the grades and retention of Hispanic students. *College Student Affairs Journal*, 14, 30–36.
- Gough, H. G. (1978). Some predictive implications of premedical scientific competence and preferences. *Journal of Medical Education*, 53(4), 291–300.
- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical statistical controversy. *Psychology, Public Policy, and Law*, 2, 293–323.
- Hanson, M. D., Dore, K. L., Reiter, H. I., & Eva, K. W. (2007). Medical school admissions: Revisiting the veracity and independence of completion of an autobiographical screening tool. *Academic Medicine*, 82, S8–S11. doi:[10.1097/ACM.0b013e3181400068](https://doi.org/10.1097/ACM.0b013e3181400068).
- Huff, K. L., & Fang, D. (1999). When are students most at risk of encountering academic difficulty? A study of the 1992 matriculants to US medical schools. *Academic Medicine*, 74(4), 454–460. doi:[10.1097/00001888-199904000-00047](https://doi.org/10.1097/00001888-199904000-00047).
- Huff, K. L., Koenig, J. A., Treptau, M. M., & Sireci, S. G. (1999). Validity of MCAT scores for predicting clerkship performance of medical students grouped by sex and ethnicity. *Academic Medicine*, 74(10, Suppl), S41–S44. doi:[10.1097/00001888-199910000-00035](https://doi.org/10.1097/00001888-199910000-00035).
- Huffcutt, A. I., & Arthur, W., Jr. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *The Journal of Applied Psychology*, 79(2), 184–190. doi:[10.1037/0021-9010.79.2.184](https://doi.org/10.1037/0021-9010.79.2.184).
- Julian, E. R. (2005). Validity of the Medical College Admission Test for predicting medical school performance. *Academic Medicine*, 80(10), 910–917. doi:[10.1097/00001888-200510000-00010](https://doi.org/10.1097/00001888-200510000-00010).
- Kirchner, G. L., & Holm, M. B. (1997). Prediction of academic and clinical performance of occupational therapy students in an entry-level master's program. *The American Journal of Occupational Therapy*, 51, 775–779.
- Kirchner, G. L., Stone, R. G., & Holm, M. B. (2000). Use of admission criteria to predict performance of students in an entry-level master's program on fieldwork placements and in academic courses. *Occupational Therapy in Health Care*, 13(1), 1. doi:[10.1300/J003v13n01_01](https://doi.org/10.1300/J003v13n01_01).
- Koenig, J. A. (1992). Comparison of medical school performances and career plans of students with broad and with science-focused premedical preparation. *Academic Medicine*, 67(3), 191–196. doi:[10.1097/00001888-199203000-00011](https://doi.org/10.1097/00001888-199203000-00011).
- Koenig, J. A., Sireci, S. G., & Wiley, A. (1998). Evaluating the predictive validity of MCAT scores across diverse applicant groups. *Academic Medicine*, 73(10), 1095–1106. doi:[10.1097/00001888-199810000-00021](https://doi.org/10.1097/00001888-199810000-00021).
- Kreiter, C. D., & 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(2), 95–100.
- Kreiter, C. D., Yin, P., Solow, C., & Brennan, R. L. (2004). Investigating the reliability of the medical school admissions interview. *Advances in Health Sciences Education: Theory and Practice*, 9(2), 147–159. doi:[10.1023/B:AHSE.0000027464.22411.0f](https://doi.org/10.1023/B:AHSE.0000027464.22411.0f).
- 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. doi:[10.1207/S15328015TLM1401_9](https://doi.org/10.1207/S15328015TLM1401_9).
- Lewis, N. J., Rees, C. E., Hudson, J. N., & Bleakley, A. (2005). Emotional intelligence in medical education: Measuring the unmeasurable? *Advances in Health Sciences Education*, 10, 339–355. doi:[10.1007/s10459-005-4861-0](https://doi.org/10.1007/s10459-005-4861-0).
- Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgement tests: A comparison in terms of predictive validity. *The Journal of Applied Psychology*, 91(5), 1181–1188. doi:[10.1037/0021-9010.91.5.1181](https://doi.org/10.1037/0021-9010.91.5.1181).
- Lievens, F., Buyse, T., & Sackett, P. R. (2005). The operational validity of a video-based situational judgement test for medical college admissions: Illustrating the importance of matching predictor and

- criterion construct domains. *The Journal of Applied Psychology*, 90(3), 442–452. doi:[10.1037/0021-9010.90.3.442](https://doi.org/10.1037/0021-9010.90.3.442).
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *The Journal of Applied Psychology*, 79, 599–616. doi:[10.1037/0021-9010.79.4.599](https://doi.org/10.1037/0021-9010.79.4.599).
- McManus, I. C., & Richards, B. (1986). Prospective study of performance of medical students during pre-clinical years. *British Medical Journal*, 293, 124–127.
- Meredith, K. E., Dunlap, M. R., & Baker, H. H. (1982). Subjective and objective admissions factors as predictors of clinical clerkship performance. *Journal of Medical Education*, 57(10 Pt 1), 743–751.
- Neame, R. L., Powis, D. A., & Bristow, T. (1992). Should medical students be selected only from recent school-leavers who have studied science? *Medical Education*, 26(6), 433–440. doi:[10.1111/j.1365-2923.1992.tb00202.x](https://doi.org/10.1111/j.1365-2923.1992.tb00202.x).
- Norcini, J. J., Lipner, R. S., & Kimball, H. R. (2002). Certifying examination performance and patient outcomes following acute myocardial infarction. *Medical Education*, 36(9), 853–859. doi:[10.1046/j.1365-2923.2002.01293.x](https://doi.org/10.1046/j.1365-2923.2002.01293.x).
- O'Callaghan, K. W., & Bryant, C. (1990). Noncognitive variables: A key to black-American academic success at a military academy. *Journal of College Student Development*, 31, 121–126.
- O'Halloran, C. M., Altmaier, E. M., Smith, W. L., & Franken, E. A., Jr. (1993). Evaluation of resident applicants by letters of recommendation: A comparison of traditional and behavior-based formats. *Investigative Radiology*, 28(3), 274–277.
- Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgement inventory as predictors of college student performance. *The Journal of Applied Psychology*, 89, 187–207. doi:[10.1037/0021-9010.89.2.187](https://doi.org/10.1037/0021-9010.89.2.187).
- Ozuah, P. O. (2002). Variability in Deans' letters. *Journal of the American Medical Association*, 288, 1061. doi:[10.1001/jama.288.9.1061](https://doi.org/10.1001/jama.288.9.1061).
- Papadakis, M. A., Teherani, A., Banach, M. A., Knettlar, T. R., Rattner, S. L., Stern, D. T., et al. (2005). Disciplinary action by medical boards and prior behavior in medical school. *The New England Journal of Medicine*, 353(25), 2673–2682. doi:[10.1056/NEJMsa052596](https://doi.org/10.1056/NEJMsa052596).
- Peskun, C., Detsky, A., & Shandling, M. (2007). Effectiveness of medical school admissions criteria in predicting residency ranking four years later. *Medical Education*, 41(1), 57–64. doi:[10.1111/j.1365-2929.2006.02647.x](https://doi.org/10.1111/j.1365-2929.2006.02647.x).
- Powis, D. A., Neame, R. L., Bristow, T., & Murphy, L. B. (1988). The objective structured interview for medical student selection. *British Medical Journal (Clinical Research Ed.)*, 296(6624), 765–768.
- Ramsey, P. G., Carline, J. D., Inui, T. S., Larson, E. B., LoGerfo, J. P., & Wenrich, M. D. (1989). Predictive validity of certification by the American Board of Internal Medicine. *Annals of Internal Medicine*, 110(9), 719–726.
- 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.
- Ross, C. A., & Leichner, P. (1984). Criteria for selecting residents: A reassessment. *Canadian Journal of Psychiatry*, 29, 681–686.
- Salvatori, P. (2001). Reliability and validity of admissions tools used to select students for the health professions. *Advances in Health Sciences Education*, 6, 159–175. doi:[10.1023/A:1011489618208](https://doi.org/10.1023/A:1011489618208).
- Schofield, W., & Garrard, J. (1975). Longitudinal study of medical students selected for admission to medical school by actuarial and committee methods. *British Journal of Medical Education*, 9(2), 86–90.
- Sedlacek, W. E. (1991). Using noncognitive variables in advising nontraditional students. *National Academic Advising Association Journal*, 11(1), 75–82.
- Sedlacek, W. E. (1996). An empirical method of determining nontraditional group status. *Measurement and Evaluation in Counseling and Development*, 28, 200–210.
- Sedlacek, W. E. (1999). Blacks on White campuses: 20 years of research. *Journal of College Student Development*, 40, 538–550.
- Sedlacek, W. E. (2003). Alternative measures in admissions and scholarship selection. *Measurement and Evaluation in Counseling and Development*, 35, 263–272.
- Sedlacek, W. E. (2004). *Beyond the big test—noncognitive assessment in higher education*. San Francisco: Jossey-Bass.
- Sedlacek, W. E., & Adams-Gaston, J. (1992). Predicting the academic success of student-athletes using SAT and noncognitive variables. *Journal of Counseling and Development*, 70(6), 724–727.
- Shapiro, R. S., Simpson, D. E., Lawrence, S. L., Talsky, A. M., Sobocinski, K. A., & Schiedermayer, D. L. (1989). A survey of sued and nonsued physicians and suing patients. *Archives of Internal Medicine*, 149, 2190–2196. doi:[10.1001/archinte.149.10.2190](https://doi.org/10.1001/archinte.149.10.2190).

- Standridge, J., Boggs, K. J., & Muga, K. A. (1997). Evaluation of the health occupations aptitude examination. *Respiratory Care*, *42*, 868–872.
- Tamblyn, R., Abrahamowicz, M., Brailovsky, C., Grand'Maison, P., Lescop, J., Norcini, J., et al. (1998). Association between licensing examination scores and resource use and quality of care in primary care practice. *Journal of the American Medical Association*, *280*(11), 989–996. doi:[10.1001/jama.280.11.989](https://doi.org/10.1001/jama.280.11.989).
- Tamblyn, R., Abrahamowicz, M., Dauphinee, W. D., Hanley, J. A., Norcini, J., Girard, N., et al. (2002). Association between licensure examination scores and practice in primary care. *Journal of the American Medical Association*, *288*(23), 3019–3026. doi:[10.1001/jama.288.23.3019](https://doi.org/10.1001/jama.288.23.3019).
- Tamblyn, R., Abrahamowicz, M., Dauphinee, D., Wenghofer, E., Jacques, A., Klass, D., et al. (2007). Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *Journal of the American Medical Association*, *298*(9), 993–1001. doi:[10.1001/jama.298.9.993](https://doi.org/10.1001/jama.298.9.993).
- Taylor, P., & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behaviour employment interview questions. *Journal of Occupational and Organizational Psychology*, *75*, 277–294. doi:[10.1348/096317902320369712](https://doi.org/10.1348/096317902320369712).
- Ting, S. R. (1997). Estimating academic success in the first year of college for specially admitted white students: A model combining cognitive and psychosocial predictors. *Journal of College Student Development*, *38*, 401–409.
- Tracey, T. J., & Sedlacek, W. E. (1985). The relationship of noncognitive variables to academic success: A longitudinal comparison by race. *Journal of College Student Personnel*, *26*, 405–410.
- Tracey, T. J., & Sedlacek, W. E. (1987). Prediction of college graduation using noncognitive variable by race. *Measurement and Evaluation in Counseling and Development*, *19*, 177–184.
- Tracey, T. J., & Sedlacek, W. E. (1988). A comparison of white and black student academic success using noncognitive variables: A LISREL analysis. *Research in Higher Education*, *27*, 333–348. doi:[10.1007/BF00991662](https://doi.org/10.1007/BF00991662).
- Tracey, T. J., & Sedlacek, W. E. (1989). Factor structure of the noncognitive questionnaire—revised across samples of black and white college students. *Educational and Psychological Measurement*, *49*, 637–648. doi:[10.1177/001316448904900316](https://doi.org/10.1177/001316448904900316).
- Trail, C., Reiter, H. I., Bridge, M., Stefanowska, P., Schmuck, M., & Norman, G. (2006). Impact of field of study, college and year on calculation of cumulative grade point average. *Advances in Health Sciences Education Theory and Practice* (Epub ahead of print).
- Violato, C., & Donnon, T. (2005). Does the medical college admission test predict clinical reasoning skills? A longitudinal study employing the Medical Council of Canada clinical reasoning examination. *Academic Medicine*, *80*(10, Suppl), S14–S16. doi:[10.1097/00001888-200510001-00007](https://doi.org/10.1097/00001888-200510001-00007).
- Wiesner, W., & Cronshaw, S. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology*, *61*, 275–290.
- Woodward, C. A., & McAuley, R. G. (1983). Can the academic background of medical graduates be detected during internship? *Canadian Medical Association Journal*, *129*(6), 567–569.
- Yens, D. P., & Stimmel, B. (1982). Science versus nonscience undergraduate studies for medical school: A study of nine classes. *Journal of Medical Education*, *57*(6), 429–435.