

# Selection of medical students on the basis of non-academic skills: is it worth the trouble?

Authors: A Susan M Niessen<sup>A</sup> and Rob R Meijer<sup>B</sup>

## ABSTRACT

In this article, we discuss the practical usefulness of selecting future medical students on the basis of increasingly popular non-academic tests (eg multiple mini-interviews, situational judgment tests) in addition to academic tests. Non-academic tests assess skills such as ethical decision making, communication and collaboration skills, or traits such as conscientiousness. Although other studies showed that performance on non-academic tests could have a positive relationship with future professional performance, we argue that this relationship should be interpreted in the context of the base rate (the proportion of suitable candidates in the applicant pool) and the selection ratio (the proportion of selected applicants from the applicant pool). We provide some numerical examples in the context of medical student selection. Finally, we suggest that optimising training in non-academic skills may be a more successful alternative than selecting students on the basis of these skills.

**KEYWORDS:** Selective admission, utility, validity, medical students

## Introduction

There is an increasing interest in the selection of future medical students on the basis of non-academic tests in addition to, or instead of, the traditional cognition-based or knowledge-based academic tests such as the Medical College Admission Test® (MCAT). Non-academic tests measure skills such as communication skills, professional behaviour, and ethical decision making, or traits such as personality characteristics. An example of a non-academic skills admission test is the multiple mini-interview (MMI).<sup>1,2</sup> The MMI consists of a series of short, structured interviews and tasks where potential students show their interpersonal skills and ethical standards. Another example is the use of (video-based) situational judgment tests (SJTs) that contain social doctor-patient and doctor-colleague interactions and where candidates have to say how they would respond to a particular situation.<sup>3,4</sup> Other non-academic admissions tests, often in the form of self-report questionnaires, measure non-cognitive traits such

as empathy.<sup>5</sup> In this article, we focus on non-academic skills. However, it is not always clear whether the constructs measured by non-academic tests should be considered skills, traits, or a combination of both. SJTs, for example, are more related to cognitive abilities when knowledge instructions are used (how should one act), and more related to personality traits when behavioural instructions are used (how would you act).<sup>6</sup>

The central idea of using these measures on top of academic measures like high school grade point average (GPA) and standardised test scores (eg MCAT scores) is that these measures improve the selection procedure of future medical students. That is, through the use of non-academic tests such as MMIs or SJTs, the candidates selected will perform better as a doctor than those who were selected on the basis of academic measures only. In 2015, Harris and colleagues<sup>7</sup> were critical about the use of 'non-knowledge-based tests and situational judgement tests to test desirable professional attitudes, such as empathy and ethical awareness'. Their main argument was that the validity of these tests to predict academic performance was low as compared with knowledge-based tests. Instead, they advocated using knowledge-based tests to select future doctors. Indeed, non-academic tests should show predictive validity and incremental validity over and above academic tests. However, using academic performance as a criterion may not be very useful for this purpose because the aim is not to predict academic performance but doctor performance or professional performance. To show incremental validity, these tests should be reliable instruments (ie test results should be consistent across replications), should show positive relationships with relevant criteria (such as professional performance), and should not be strongly related to academic tests. Very few studies address incremental validity of non-academic tests over and above academic tests using such criteria (Lievens<sup>3</sup> and Adam *et al*<sup>5</sup> are exceptions). However, statistically significant incremental validity is not necessarily equal to practically relevant incremental validity. In this article, we go beyond incremental validity and discuss this topic from a utility approach that originates from the personnel selection literature. The advantage of this approach is that it immediately shows the practical effect of using different selection instruments. Our main message is that in many medical student selection situations the incremental validity and utility of the use of non-academic instruments seems to be small and that the recent trend to use these instruments needs a more solid empirical basis.

**Authors:** <sup>A</sup>PhD candidate, Department of Psychometrics and Statistics, University of Groningen, Groningen, Netherlands; <sup>B</sup>professor of psychometrics, University of Groningen, Groningen, Netherlands

### Practical usefulness of additional selection instruments

The effects of additional predictors above academic tests to select students will be largest when the correlation between predictors is low. However, several studies have shown that non-academic skills and academic, cognition-based measures are not independent. For example, in a recent study on medical student selection from the Netherlands<sup>8</sup> it was concluded that

*Top pre-university GPA students also achieved the highest possible score in the professionalism course most often. In this course, non-academic variables such as interpersonal and communication skills, ethical decision-making, reflection and professional behaviour are assessed. The overall high performance of the top pre-university GPA group suggests that applicants who perform well academically might also have an advantage in the so-called non-academic domain.*

Thus, in this course, students were judged on showing professional behaviour and their reflections on professional behaviour.

These findings are also in line with a study from the USA in which it was concluded that ‘the relationships between cognitive and non-cognitive subdomains of the licensing examinations reported here ranged between  $r=0.17$  and  $r=0.43$  and correlations generally increased with trainees’ seniority’.<sup>2</sup> This increase with trainees’ seniority also reflected that non-academic skills are, to a large extent, trainable. Note that the non-cognitive skills mentioned in this study are similar to what we consider non-academic skills. In a more general (non-medical) context, a meta-analysis found moderate relationships between SJT scores and cognitive ability, with a stronger relationship for tests that were based on job analysis and tests that provided knowledge instructions.<sup>6</sup>

Thus, the studies cited above showed the same pattern; namely that academic and non-academic skills are positively correlated. However, non-academic skills may still improve predictions based on only academic scores. For example, in a Belgian study<sup>3</sup> the additional value of a video-based SJT that measured interpersonal skills in addition to knowledge test scores was investigated. Lievens<sup>3</sup> investigated the additional explained variance above high school GPA and knowledge test scores in a hierarchical regression analysis and found a significant effect. The SJT had significant added value for predicting four outcomes: interpersonal GPA, an interpersonal skills assessment, doctor performance, and performance on a case-based panel interview, with additional portions of explained variance of 4.4%, 1.4%, 2.2%, and 3.4%, respectively. On the basis of these numbers it was concluded that ‘video-based SJTs as measures of procedural knowledge about interpersonal behaviour show promise as complements to cognitive examination components’. These percentages are in agreement with a meta-analysis<sup>6</sup> that found incremental validities of SJTs over and above cognitive ability of 3–5% with job performance as a criterion. However, what do these numbers mean in terms of practical relevance?

To evaluate the practical usefulness of a selection procedure we should not only take the correlations between the predictors and the criterion scores (predictive validity) into account, but also the selection ratio (the percentage of future students who are selected), and the base rate (the percentage

of future students who would be successful without using tests).<sup>8</sup> In particular, the base rate is interesting here: what is the percentage of students who have or can obtain sufficient skills to be a successful doctor during the study without selection. Assume that we do not select candidates, how many applicants will then show sufficient skills when they get their licensing degree? There are models that show the interplay between selection ratio, base rate, and the relationship between predictor and criterion scores. The most popular model is the Taylor-Russell model,<sup>9</sup> which provides the success ratio – the proportion of admitted medical students who will perform well as doctors, for a given base rate, selection ratio, and predictive validity. This model allows us to compare the result of a selection procedure to the base rate (ie when there is no selection), or to compare the success ratios under different selection procedures. The general message of this model is that when the base rate is high or the selection ratio is high, the effect of selection is low. This is easy to understand: if most candidates will be successful anyway, or almost all candidates will be selected, selection will have only a small effect on the quality of the admitted students, regardless of the predictive value of the selection instruments. Instead of consulting the Taylor-Russell tables, calculations can also be made using a web-application.<sup>10</sup>

### An illustration: Dutch medical selection

How does this translate to the selection of future medical students? To answer this question, we need information from the selection ratio and the base rate. In general, medical schools are selective, and thus the selection ratio will be moderate to low. In the USA, for example, selection ratios range from 0.02 for some very selective medical schools to approximately 0.60. In the Netherlands, the selection ratio is around 0.60.<sup>11</sup> The base rate is around 0.80 for successful doctor performance. This base rate is high for two reasons:

- 1 Students who apply to study medicine are not a random sample from the general population but are already strongly preselected on the basis of academic and non-academic skills as a result of high school selection and training.
- 2 Students are trained in academic and non-academic skills during medical education.

As a result of the interplay between base rate, selection ratio, and (multiple) correlation between predictor scores and criterion scores, we can determine the success ratio. The question is what the difference in the success ratio is when we use an academic test and a non-academic test, in contrast with using an academic test alone. In Table 1, we show these percentages for three hypothetical base rates: 0.70, 0.80, and 0.90, and four hypothetical selection ratios: 0.60, 0.30, 0.10, and 0.05. These selection ratios represent roughly the Dutch context: strict selection, like in France, where after the first year of medical school, approximately 20% of the students are selected for the second year, and very strict selection as often found in the USA and the UK. For this illustration we used the results from the Belgian study<sup>3</sup> since it was the only study that reported incremental validity of a non-academic test over and above academic tests, and used suitable criterion measures. In that study,<sup>3</sup> knowledge-based tests showed a correlation of  $r=0.07$  with doctor performance. Adding an SJT resulted in an

**Table 1. Success ratio as a function of the base rate (BR), selection ratio (SR) and low predictor-criterion correlation.**

SR	BR=0.70			BR=0.80			BR=0.90		
	R=0.07	R=0.17	Δ	R=0.07	R=0.17	Δ	R=0.07	R=0.17	Δ
0.60	71.6	73.9	2.3	81.3	83.1	1.8	90.8	92.0	1.2
0.30	72.8	76.7	3.9	82.2	85.3	3.1	91.4	93.3	1.9
0.10	74.2	79.8	5.6	83.3	87.6	4.3	92.0	94.5	2.5
0.05	74.9	81.2	6.3	83.8	88.6	4.8	92.3	95.1	2.8

$\Delta R^2 = (0.17)^2 - (0.07)^2 = 0.0289 - 0.0049 = 0.024$ ;  $\Delta$  denotes the differences in more successful decisions when the multiple correlation increase from  $R=0.07$  to  $R=0.17$  (or equivalently  $\Delta R^2 = 0.024$ ).

increase in explained variance of  $\Delta R^2 = 0.024$ , yielding a multiple validity coefficient of  $R = 0.17$ . Note that this correlation is much lower than the often reported correlations between academic test scores and academic performance in medical school. For example, in a study from the USA<sup>12</sup> median correlations of  $r = 0.46$  were reported between MCAT scores and mean grades over the first 3 years. The low correlation is probably due to the more heterogeneous criterion of doctor performance as compared with the criterion of school grades, although in the Belgian study,<sup>3</sup> the definition of 'doctor performance' was unclear. The only information that was provided was that it was based on supervisor ratings for GPs. To investigate the utility of these findings, we determined the success ratios when the (multiple) validity coefficient increased from 0.07 to 0.17.

In Table 1, it can be seen that for a selection ratio of 0.60 and a base rate of 0.80, the selection procedure with only an academic test will result in an additional 1.3% successful students compared with the base rate. Now, suppose that through the use of an additional non-academic test the predictive validity (multiple correlation) increases from 0.07 to 0.17. The reader can see in Table 1 that for a base rate of 0.80 and a selection ratio of 0.60 the percentage of successful students increases by 3.1%. However, note that the difference is only 1.8% (1.3% versus 3.1%). For a base rate of 0.90 this percentage is around 1.2% and for a base rate of 0.70 it is approximately 2.3%. Even for a selection ratio of 0.30, an increase in  $R^2$  to 0.024 only leads to a difference of 3.1% using a base rate of 0.80, 1.9% for a base rate of 0.90 and 3.9% for a base rate of 0.70. Thus, when we assume that 80% of medical school applicants would be successful in their medical job, using this SJT only provides a modest increase in the percentage of successful doctors. Only when selection is strict and base rates are not very high, using this additional non-academic test yields a larger, but still modest, increase in successful doctors (4.8% for a base rate of 0.80 and a strict selection ratio of 0.05).

Given that tests like MMIs and SJTs are complex instruments to develop and do require extensive resources, one may ask whether these extra resources pay off in practical student selection. For example, let us consider the selection of future medical students in the Netherlands. In 2015, there were 2,785 places available in eight different Dutch medical schools. With a base rate of 0.80 and a selection ratio of 0.60, this would yield 2,264 successful doctors with only academic selection, and 2,314 successful doctors with academic and non-academic selection. The difference between selection with an academic test only and selection using both tests is 50 more successful doctors when

both tests are used. In the Netherlands, these 50 extra successes are divided between eight medical schools, so on average this will result in approximately six extra successes per school. This number is small, whereas the costs of selection are often not trivial. Each medical school may take these numbers and costs into account when deciding to use non-academic tests.

A counterargument is that, although there may be few future medical students who are unsuited for the medical profession, it is still very important to detect these few students because they may bring serious harm to patients and to the medical institution. This is true, but we should show how well instruments like non-academic tests meet this goal based on analyses as shown earlier. Following our example, let us assume that 20% of the medical school applicants are not suitable to be doctors. Using an SJT in addition to an academic test would reduce the number of unsuitable doctors by between 1.8% and 4.8%, depending on the selection ratio.

There are two other arguments why institutions, in our view, should take calculations like the ones given above into account in selecting students on the basis of non-academic skills. The first is that the environment also determines professional behaviour.<sup>13</sup> The second argument is that many (or at least some) non-academic skills are trainable<sup>14</sup> and that experience is very important. As Henry Marsh<sup>15</sup> recently remarked: 'surgery is a practical craft, and you learn it largely by doing it – simulators and training courses can take you only part of the distance'. In fact, a reasonable question is: what is more effective? Selecting students based on non-academic skills, or improving the curriculum and working conditions with respect to teaching and developing these skills?

## Discussion

Medical school admissions need not be based on predicted school performance or even in improved medical job performance, defined more broadly. Medical schools can opt to select future medical students on the basis of particular talents (eg leadership, wanting to work in third world countries, etc) or on the basis of school diversity. For example, some schools use standardised test scores as diagnostic tools to predict which students need extra help in their studies.<sup>12</sup> However, it is important to realise that when predictors are chosen on the basis of the strength of the relationship with a criterion, base rate and selection ratio play a crucial role when we determine the practical outcome. Because instruments like SJTs and MMIs are psychometrically more complex, expensive to develop

and administer, and sometimes less reliable than academic instruments,<sup>6</sup> the added value may be modest and the costs high. We illustrated this in the context of medical selection but this does not show a general result in terms of the utility of non-academic tests. It is true that selection ratio and base rate may differ across medical schools and across countries. So there may be situations where it pays off to add non-academic skills tests to academic tests. Besides utility models that show increments in success ratios, there are utility models that can be used to estimate the economic consequences of implementing selection instruments.<sup>16–18</sup> However, it is not easy to illustrate these models given the uncertainty of how to estimate some parameters in these models, such as the standard deviation of economic gain from increased performance (Holling<sup>19</sup> provides some solutions). In addition, the costs of educating a student who is unsuitable for medical practice is also difficult to determine and will vary considerably across countries and universities.<sup>20</sup> A utility analysis is complex and beyond the scope of this review, but we encourage decision makers to explore the applications of these models in their particular selection context.

Another argument for selecting future students based on non-academic skills is that it may result in self-selection of future students, leading to applicants who perhaps fit the profile of a future successful doctor better. If this would be the case, then the base rate would change as a result of the selection procedure. This is not taken into account in the Taylor-Russell model discussed previously and this hypothesis deserves future research.

We would like to stress that we are not against the selection of future medical students on the basis of non-academic skills, but we would argue for future research that is aimed at the incremental validity and utility of these instruments in the context of base rates and selection ratio. As we showed, statistically significant incremental validity is not necessarily practically relevant. In future research, much more attention should be given to the criterion variables. Doctor performance is a complex variable that is not taken into account in many studies or is not operationalised clearly. For example, in the Belgian study,<sup>3</sup> students following a career in general practice were studied. It may be the case that non-academic skills like social skills are more important for this specialty than for other medical specialties. ■

### Conflicts of interest

The authors declare no conflicts of interest.

### References

- 1 Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Med Educ* 2004;38:314–26.
- 2 Eva KW, Reiter HI, Trinh K *et al*. Predictive validity of the multiple mini-interview for selecting medical trainees. *Med Educ* 2009;43:767–75.
- 3 Lievens F. Adjusting medical school admission: assessing interpersonal skills using situational judgment tests. *Med Educ* 2013;47:182–9.
- 4 Patterson F, Baron H, Carr V, Plint Lane P. Evaluation of three short-listing methodologies for selection into postgraduate training in general practice. *Med Educ* 2009;43:50–7.
- 5 Adam J, Bore M, Childs R *et al*. Predictors of professional behaviour and academic outcomes in a UK medical school: a longitudinal cohort study. *Med Teach* 2015; 37:868–80.
- 6 McDaniel MA, Hartman NS, Whetzel DL, Grubb WL. Situational judgment tests, response instructions, and validity: a meta-analysis. *Pers Psychol* 2007;60:63–91.
- 7 Harris BHL, Walsh JL, Lammey L. UK medical selection: lottery or meritocracy. *Clin Med* 2015;1:40–6.
- 8 Schripsema NR, van Trigt AM, Borleffs JCC, Cohen-Schotanus J. Selection and study performance: comparing three admission processes within one medical school. *Med Educ* 2014;48:1201–10.
- 9 Taylor HC, Russell JT. The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *J Appl Psychol* 1939;23:565–78.
- 10 McLellan RA. Theoretical expectancy calculator. Available online at [www.hr-software.net/cgi/TheoreticalExpectancy.cgi](http://www.hr-software.net/cgi/TheoreticalExpectancy.cgi) [Accessed 9 May 2016].
- 11 Niessen ASM, Meijer RR. Wanneer heeft selectie in het hoger onderwijs zin? De stand van zaken anno 2015 [When does selection in higher education pay off? The situation in 2015]. *Tijdschrift voor Hoger Onderwijs* 2015;33:4–19.
- 12 Julian ER. Validity of the medical school college admission test for predicting medical school performance. *Acad Med* 2005;80:910–7.
- 13 Johnson G. The essential impact of context of organization behavior. *Acad Manage Rev* 2006;31:366–408.
- 14 Hook KM, Pfeiffer CA. Impact of a new curriculum on medical students' interpersonal and interviewing skills. *Med Educ* 2007;41:154–9.
- 15 Marsh H. *The biggest victim of a doctors' strike would be trust in the medical profession*. London: The Guardian, 2015. Available online at [www.theguardian.com/commentisfree/2015/oct/21/doctors-strike-medical-profession-nhs](http://www.theguardian.com/commentisfree/2015/oct/21/doctors-strike-medical-profession-nhs) [Accessed 9 May 2016].
- 16 Cascio WF. Responding to the demand for accountability: a critical analysis of three utility models. *Organ Behav Hum Perform* 1980;25:32–45.
- 17 Brogden HE. When testing pays off. *Pers Psychol* 1949;2:171–85.
- 18 Cronbach LJ, Gleser GC. *Psychological tests and personnel decisions*, 2nd edn. Urbana, Illinois: University of Illinois Press, 1965.
- 19 Holling H. Utility analysis of personnel selection an overview and empirical study based on objective performance measures. *Meth Psychol Res* 1989;3:5–24.
- 20 Mellenbergh GJ. Selectie aan de poort [Selection at the gates]. *Spiegeloo* 1995;1:8–9.

**Address for correspondence:** Ms ASM Niessen, Department of Psychometrics and Statistics, Faculty of Behavioral and Social Sciences, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands.  
**Email:** [a.s.m.niessen@rug.nl](mailto:a.s.m.niessen@rug.nl)