Do personality traits assessed on medical school admission predict exit performance? A UK-wide longitudinal cohort study

R. K. MacKenzie[1] · J. Dowell[2] · D. Ayansina[3] ·
J. A. Cleland[1]

Abstract Traditional methods of assessing personality traits in medical school selection have been heavily criticised. To address this at the point of selection, "non-cognitive" tests were included in the UK Clinical Aptitude Test, the most widely-used aptitude test in UK medical education (UKCAT: http://www.ukcat.ac.uk/). We examined the predictive validity of these non-cognitive traits with performance during and on exit from medical school. We sampled all students graduating in 2013 from the 30 UKCAT consortium medical schools. Analysis included: candidate demographics, UKCAT non-cognitive scores, medical school performance data—the Educational Performance Measure (EPM) and national exit situational judgement test (SJT) outcomes. We examined the relationships between these variables and SJT and EPM scores. Multilevel modelling was used to assess the relationships adjusting for confounders. The 3343 students who had taken the UKCAT non-cognitive tests and had both EPM and SJT data were entered into the analysis. There were four types of non-cognitive test: (1) libertariancommunitarian, (2) NACE—narcissism, aloofness, confidence and empathy, (3) MEARS—self-esteem, optimism, control, self-discipline, emotional-nondefensiveness (END) and faking, (4) an abridged version of 1 and 2 combined. Multilevel regression showed that, after correcting for demographic factors, END predicted SJT and EPM decile. Aloofness and empathy in NACE were predictive of SJT score. This is the first large-scale study examining the relationship between performance on non-cognitive selection tests and medical school exit assessments. The predictive validity of these tests was limited, and the relationships revealed do not fit neatly with theoretical expectations. This study does not support their use in selection.

✉ R. K. MacKenzie
r.k.mackenzie@abdn.ac.uk

[1] Senior Clinical Lecturer in Medical Education, Institute of Education for Medical and Dental Sciences, University of Aberdeen, Room 2:036, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK

[2] Dundee Medical School, University of Dundee, Dundee, UK

[3] Medical Statistics Team, College of Life Sciences and Medicine, University of Aberdeen, Aberdeen, UK

Keywords Medical school admissions · Medical school selection · Non-cognitive testing · Psychometric testing · Situational judgement tests · United Kingdom clinical aptitude test (UKCAT)

Abbreviations

END	Emotional non-defensiveness (a domain within MEARS)
EPM	Educational performance measure (a measure of examination performance during medical school)
EU	European Union
IMD	Index of multiple deprivation, a socio-economic indicator, based on postcode (quintiles)
ITQ	Interpersonal traits questionnaire (see also NACE)
IVQ	Interpersonal values questionnaire
MEARS	Managing Emotions and Resiliency Scale
NACE	A psychometric test with the domains of narcissism, aloofness, confidence and empathy (see also ITQ)
NSSEC	National statistics socio-economic classification, based on parental occupation (quintiles)
PQA	Personal qualities assessment (includes the IVQ and ITQ)
SEC	Socio-economic classification
SJT	Situational judgement test
UCAS	Universities and colleges admissions service, an organisation whose main role is to operate the application process to British Universities
UKCAT	United Kingdom clinical aptitude test
UKFPO	United Kingdom foundation programme office

Introduction

There are a number of issues of importance in selection for admission to medical school (Prideaux et al. 2011; Girotti et al. 2015). One of these is assessing the predictive validity and reliability of any selection tool, to ensure it measures what it claims to measure, does so fairly and consistently and can be employed rationally (e.g., Cleland et al. 2012; Norman 2015). A second is ensuring that selection tools assess the range of attributes considered important by key stakeholders. Medical schools must select applicants who will not only excel academically but also possess personality traits befitting a career in medicine such as compassion, team working skills and integrity (e.g., Albanese et al. 2003; General Medical Council 2009; Frank and Snell 2015; Accreditation Council for Graduate Medical Education 2014).

This increasing recognition that there is more to being a capable medical student or doctor than academic performance follows on from a similar direction of travel in education where, according to a large body of research, a number of non-cognitive skills are associated with positive academic and work-related outcomes for young people (see Gutman and Schoon 2013, for a recent review). Given this, the assessment of non-academic factors, or personality traits is of increasing importance in medical school selection (Patterson 2013). However, "traditional" methods of assessing such personality traits, including unstructured interviews, using personal references and autobiographical statements are now known to have weak predictive validity (Cleland et al. 2012; Patterson et al.

2016). There is a drive to identify better ways to assess non-academic attributes such as values and personality traits in medical school selection.

Various different ways to do so have been proposed. These can be grouped into "paper and pencil" assessments of personality traits (e.g., Adams et al. 2012, 2015; Bore et al. 2005a, b; Dowell et al. 2011; Fukui et al. 2014; James et al. 2013; Lumsden et al. 2005; Manuel et al. 2005; Nedjat et al. 2013), structured multiple interview approaches (Dore et al. 2010; Eva et al., 2004a, b, 2009; Hofmeister et al. 2008, 2009; O'Brien et al. 2011; Reiter et al. 2007; Roberts et al. 2008; Rosenfeld et al. 2008), selection centres (Gafni et al. 2012; ten Cate and Smal 2002; Ziv et al. 2008; Gale et al. 2010; Randall et al. 2006a, b) and—the "new kid on the block"—situational judgement tests (Christian et al. 2010; Koczwara et al. 2012; Lievens 2013; Lievens et al. 2008; Patterson et al. 2009).

However, there are relatively few studies examining the predictive validity of the "paper and pencil" tests which aim to assess personality traits in medical school applicants. Those which have been published are often concerned with feasibility of use across cultural settings (e.g., Fukui et al. 2014; Nedjat et al. 2013) and/or are descriptive in terms of cross-sectional comparisons across different groups of students (e.g., graduate entrants versus school-leavers: Bore et al. 2005a, b; James et al. 2013; Lumsden et al. 2005; Nedjat et al. 2013). The few studies of predictive validity to date tend to be small scale, usually single site (Adams et al. 2012, 2015; Manuel et al. 2005) and/or use local assessments as outcome measures (Adams et al. 2012, 2015; Dowell et al. 2011; Manuel et al. 2005), limiting their generalizable messages. Large-scale, independent studies of the predictive validity of approaches to assessing personality traits, or non-academic factors in medical selection are lacking, partly because appropriate non-academic outcome markers are not easily available.

Moreover, there is much debate about the promise of personality traits for predicting success generally, the different approaches being advocated to measure these, and a clear need for more evidence (Norman 2015; Powis 2015). Drawing on the wider educational literature again, it is clear that personality traits include a very broad range of characteristics. These can be separated into those considered to be modifiable, such as motivation, resilience, perseverance, and social and communication skills, and those considered more stable or personality traits, which include Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (also called Emotional Stability) (Gutman and Schoon 2013). There is a wealth of evidence indicating that the latter, the "Big Five" personality traits, correlate highly with job performance over a range of occupational groups (e.g., Barrick and Mount 1991; Rothmann and Coetzer 2003; Salgado 1997; Dudley et al. 2006) and with performance at medical school (e.g., Lievens et al. 2002). It is this apparently stable group of traits which has been used as the theoretical basis of most "paper and pencil" assessments of personality traits designed specifically for use in medical school selection (see earlier for references). However, the more recent approaches to measuring personal characteristics in medical school selection have a slightly different conceptual basis. For example, rather than being based directly on the "Big Five" theory of personality traits, SJTs are based on implicit trait policy (ITP) theory and, depending on the job level, specific job knowledge (e.g., Motowidlo and Beier 2010a, b; Patterson et al. 2015a, b). SJTs measure the expression of personality traits in hypothetical situations which are designed on the basis of what is expected in the job for which the individual is being assessed (Motowidlo et al. 2006). They encompass measurement of personal choice (e.g., what is the best way to respond in this particular situation?) rather than just unfiltered (our word) trait expression which is arguably what is measured in traditional personality tests. There is also a pragmatic difference between "paper and pencil" tests and the SJTs.

The latter are based on thorough job analysis (Patterson et al. 2012b, Motowidlo et al. 1990) of what is expected by doctors in particular roles (e.g., junior doctor (resident)/or doctor working in a particular specialty) and take the stance that "one size does not fit all", whereas the former are typically more general measures of traits which are considered generically important to being a doctor. We return to the implications of these different positions and theoretical underpinnings for assessing personality traits in medical school section in the discussion section of this paper.

Several major changes in selection for medical school and medical training after graduation in the UK now enable large-scale multi-site studies examining the predictive validity of selection processes, including those proposing to measure personality traits. The first of these is greater consistency across UK medical schools in terms of their selection approaches (Cleland et al. 2014), with, for example, the vast majority of UK medical schools using the same aptitude test, the UK Clinical Aptitude Test (UKCAT), as part of their selection matrix. While the focus of the UKCAT is assessment of cognitive ability, "non-cognitive" or personality trait tests were included, on a trial basis, in 2007–2009. The second is the introduction of a standardised, national process for selection into the next stage of medical training after medical school in the UK, via the Foundation Programme Office (UKFPO). Those entering the selection process for the UKFPO obtain two indicators of performance: an Educational Performance Measure (EPM) and the score they achieve for a Situational Judgement Test (SJT). We present details of these indicators later in this paper. Finally, there is a move within the UK for organisations such as UKCAT and the UKFPO to work together in terms of data linkage, to enable large-scale, high-quality, longitudinal research projects.

Together, these innovations finally provide the opportunity to address a gap in the literature highlighted many years ago (see Schuwirth and Cantillon 2005). Our aim in this paper is therefore to examine the predictive power of tests purporting to assess personal personality traits in relation to two national performance indicators on exit from medical school: an academic progress measure and a measurement of personality traits determined, through a job analysis, to be associated with successful performance as a Foundation Programme doctor. We do so with data from a large number of medical schools.

Methods

Design

This was a quantitative study grounded in post-positivist research philosophy (Savin Badin and Major, Savin Baden and Major 2013). We examined the predictive validity of the personality traits, or "non-cognitive" component of the UKCAT admissions test (http://www.ukcat.ac.uk/) compared to the UK Foundation Programme (UKFPO: (http://www.foundationprogramme.nhs.uk/pages/home)) performance indicators in one graduating student cohort.

Study population

Our sample was the 2013 graduating cohort of UK medical students from the 30 UKCAT medical schools. This was the first cohort for whom both UKCAT and UKFPO indicators were available.

Data description

With appropriate permissions in place, working within a data safe haven (to ensure adherence to the highest standards of security, governance, and confidentiality when storing, handling and analysing identifiable data), routine data held by UKCAT and UKFPO were matched and linked.

The following demographic and pre-entry scores were collected: age on admission to medical school; gender; ethnicity; type of secondary school attended (fee-paying or non fee-paying); indicators of socio-economic status or classification (SEC), including Index of Multiple Deprivation (IMD) which is based on postcode, and the National Statistics Socio-economic Classification (NSSEC) which is based on parental occupation; domicile (UK, European Union [EU] or overseas); and academic achievement prior to admission, out of a maximum of 360 points (the UCAS tariff). UKCAT cognitive scores were not included given the research question focused on the predictive validity of UKCAT non-cognitive scores. Those taking the test in 2007 and 2008 were randomly allocated to sit one of four non-cognitive tests (see below). These were:

1. The Interpersonal Values Questionnaire (IVQ) measures the extent to which the respondent favours individual freedoms (versus societal rules) as a basis for making moral decisions (Bore et al. 2005a, b; Powis et al. 2005). The rationale being that this dimension of moral orientation, the extent to which the individual will 'act in own best interests' (Libertarian) vs 'act in interests of society (Communitarian). This has one domain entitled libertarian (low score –communitarian (high score). Candidates are presented with a number of situations where people have to decide what to do according to their opinions or values, responding via a 4 point Likert scale to decide where best their values sit.

2. The Interpersonal Traits Questionnaire (ITQ) or NACE, which measures narcissism, aloofness, confidence (in dealing with people) and empathy Munro et al. (2005); Powis et al. (2005). It claims to assess specific aspects of the wider domain of empathy; a high degree of empathy is linked to convivial interpersonal relationships and is generally seen as a positive thing in care-givers; although too high a degree of empathy it is argued could lead to over-involvement and burnout. ITQ produces a summary score for INVOLVEMENT where $C + E - (N + A)$, therefore some totals may be negative overall representing 'detachment)'. Overall confidence and empathy are deemed positive, narcissism and aloofness negative. The candidates who receive this test are presented with 100 statements about people and the way in which they might think or behave in certain situations. They are then given a 4 point Likert scale, and asked to decide which statements most relate to them.

3. The Managing Emotions and Resiliency Scale (MEARS) (Childs et al. 2008) was designed to reflect the cognitive, behavioural and emotional elements of resilience and describe coping styles in terms of attitudes, beliefs and typical behaviour, in six domains: self-esteem, optimism, self-discipline, faking, emotional non-defensiveness, and control (Childs 2012). In each a high score reflects a high perceived self-value in that domain. It is reported as three scores: cognitive/self-esteem and optimism scales, behavioural/control and self-discipline and emotional non-defensiveness. Candidates receive a set of paired statements that represent opposing viewpoints. They must decide their level of agreement within a six point range.

4. 1 and 2 above combined, both in an abridged format.

Note that UKCAT introduced the ITQ, IVQ and MEARS assessment on a pilot basis and the scores were not made available to selectors, i.e. NOT used in the actual selection to medical school. See Appendix: UKCAT non-cognitive test example questions.

The four outcome measures were the UKFPO selection SJT and EPM (decile and total) and total UKFPO. The EPM is a decile ranking (within each medical school) of an individual student's academic performance across all years of medical school except final year, plus additional points for extra degrees, publications etc. The total EPM score is based on three components, with a combined score of up to 50 points:

- Medical school performance by decile (presented as 34–43 points).
- Additional degrees, including intercalation (up to 5 points).
- Publications (up to 2 points).

We chose the EPM as an outcome measure as the wider education literature strongly indicates that personality traits relate to performance on academic outcomes (Gutman and Schoon 2013).The UKFPO SJT is also scored out of 50 points. The SJT focuses on key non-academic criteria deemed important for junior doctors on the basis of a detailed job analysis (Commitment to Professionalism, Coping with Pressure, Communication, Patient Focus, Effective Teamwork; see e.g., Patterson and Ashworth 2011; Patterson et al. 2015a). It presents candidates with hypothetical and challenging situations that they might encounter at work, and may involve working with others as part of a team, interacting with others, and dealing with workplace problems. In response to each situation, candidates are presented with several possible actions (in multiple choice format) that could be taken when dealing with the problem described. It is administered to all final year medical students in the UK as part of the foundation programme application process, is taken in exam conditions, and consists of 70 questions in 2 h 20 min (http://www.foundationprogramme.nhs.uk/pages/medical-students/SJT-EPM). It is a relatively new assessment but a preliminary validation study (Patterson et al. 2015a, b) has identified that that higher SJT scores were associated with higher ratings of Foundation Year 1 doctors (FY1 s: those in their first year post-graduation) in-practice performance as measured by supervisor ratings and other key performance outcomes (via supervisor ratings and routine measures); that the two selection tools (SJT and EPM) were complementary in providing prediction of performance, and that FY1 doctors in the low scoring SJT category were almost five times more likely to receive remediation than those who were in the high scoring category.

We chose this as an outcome measure as, given that there is emerging consensus that the SJT is essentially a measurement technique that targets non-cognitive attributes (Motowidlo and Beier 2010a, b), this offers a meaningful interim outcome marker for non-academic measures used within medical school selection processes.

The EPM and SJT are summed to give the UKFPO score out of 100.

Statistical analysis

All data were analysed using SPSS 22.0. Pearson or Spearman's rank correlation coefficients were used to examine the linear relationship between each of SJT score and EPM and continuous factors such as UKCAT scores and pre-admission academic scores and age. In terms of practical interpretation of the magnitude of a correlation coefficient, we have a priori defined low/weak correlation as $r = 0.10$–0.29, moderate correlation as $r = 0.30$–0.49 and strong correlation as $r \geq 0.50$. Two-sample t-tests, ANOVA, Kruskal–Wallis or Mann–Whitney U tests were used to compare UKFPO indices across levels of categorical factors as appropriate.

Multilevel linear models were constructed to assess the relationship between the independent variables of interest: UKCAT non-cognitive test totals and individual domains with each of the four outcomes (SJT, EPM decile, EPM total and UKFPO total). Fixed effects models were fitted first and then random intercepts and slopes were introduced using maximum likelihood methods. Intercepts and slopes for the medical schools were allowed to vary for the non-cognitive tests variables only. Models were adjusted for identified confounders (based on pre-hoc testing showing a correlation coefficient of >0.2 or <−0.2) such as gender, age at admission, IMD quintiles, year UKCAT exam was taken and whether or not the student attended a fee-paying school (NS-SEC and ethnicity had to be dropped from the models due to issues with non-convergence). Interactions between our primary variables and year of UKCAT exam were tested using Wald statistics and was dropped from the models if not significant at the 5 % level. Nested models were compared using information criteria such as the log-likelihood statistic, Akaike's information criteria, and Schwarz's Bayesian information criteria. The best fitting models are presented.

Results

There were 6294 students from 30 medical schools in the graduating 2013 cohort. UKCAT non-cognitive and UKFPO results were available for the 3343 students who sat the UKCAT in 2007 (n = 2714) and 2008 (n = 629) but not those who sat the test in 2006 as non-cognitive tests were not part of UKCAT in 2006—i.e. those applying in 2006 had not had the non-cognitive tests administered.

Table 1 shows the demographic profile of the cohort. Most students were from the UK (n = 2958, 90.3 %). The majority (58 %) were female and Caucasian (73.6 %). Just under a quarter of students had attended a fee paying school (23.9 %). The majority of graduating medical students were from higher SEC groups.

In terms of outcome measures, as would be expected in a decile system such as the EPM, the percentage of graduating students within each decile per school were relatively constant (varying between 9.7 and 11.2) with only the lowest decile as an outlier (7.6). EPM, SJT and total UKFPO scores are shown in Table 1. Almost one half (47.8 %) of the sample had no additional EPM points, 34.9 % (n = 1168) gained three or more further degree points, which indicates they had either intercalated or entered medicine as an Honours graduate. Most (75.3 %) did not gain any points for publications, while 18.4 % gained 1 point, and 6.3 % 2 points.

Table 2 provides an overview of candidate performance on the UKCAT non-cognitive tests. Note that each candidate sat only one of the four tests. The table shows the possible score for each domain and the range achieved by candidates (important for contextualising the multivariate analysis below) as well as the mean score and standard deviation or median and interquartile range, depending on distribution.

Table 3 shows the relationship between demographic characteristics and outcomes. Being older at the time of admission to medical school had weak positive correlation with EPM (r = 0.126, $p < 0.001$) and a weak negative correlation with SJT (r = −0.054, $p < 0.001$). Females performed significantly better than males in their EPM decile [median (IQR)]: females 6 (4, 8) versus males 5 (3, 8) $p < 0.001$. Females also had higher marks in the SJT: females 41.3 (39.1, 43.3) versus males 40.4 (38.3, 42.3) $p < 0.001$. Females outperformed males in their UKFPO scores: 82.4 (78.3, 86.4) versus males 81.1 (76.9, 85.2) $p < 0.001$. Caucasian students performed better than non-Caucasians in all outcomes

Table 1 Descriptive statistics of the demographic variables of the sample

Demographic	
All, n	3343
Age at admission (years), median (IQR)	19 (18, 22)
Missing	97
Female, n (%)	1943 (58.1)
Ethnic group, n (%)	
Caucasian	2420 (73.6)
Non-caucasian	868 (26.4)
Missing	56
Type of secondary school, n (%)	
Fee-paying	721 (23.9)
Non-Fee paying	2294 (76.1)
Missing	329
IMD quintile, n (%)	
1	1191 (39.7)
2	740 (24.6)
3	521 (17.3)
4	348 (11.6)
5	203 (6.8)
Non-UK	318
Missing	23
NS-SEC score, n (%)	
1	2517 (86.8)
2	131 (4.5)
3	151 (5.2)
4	40 (1.4)
5	60 (2.1)
Missing	445
Domicile, n (%)	
United Kingdom	2958 (90.3)
European Union	88 (2.7)
International	230 (7.0)
Missing	68

($p < 0.001$). In terms of type of secondary school attended, students who had attended independent secondary schools had a poorer EPM decile median of 6 (IQR 3, 8) than students from non-fee-paying schools (median 6, IQR 3, 8: $p < 0.001$). No statistically significant difference was seen in the other outcome measures. Spearman's rho identified a weak correlation between pre-admission academic scores and each of EPM decile ($r = 0.198$, $p < 0.001$), total EPM ($r = 0.224$, $p < 0.001$), SJT ($r = 0.104$, $p < 0.001$) and total UKFPO ($r = 0.212$. $p < 0.001$).

Linear regression showed that there was no significant association between EPM decile or total EPM and any of the individual domains in the non-cognitive tests 1, 2 and 4. In test 3, however, there was modest correlation between total EPM and each of the individual MEARS domains ($r = 0.255–0.449$, $p < 0.001$) and there was weak correlation between

Table 2 UKCAT non-cognitive domain scores

UKCAT non-cognitive domain scores		Points available and (actual results range)
Test 1, n = 879	Mean (SD)	
Libertarian communitarian, mean (SD)	118 (13.0)	45–180 (80)
Test 2 NACE, n = 973	Mean (SD)	
Narcissism	52.8 (8.2)	20–100 (56)
Aloofness	44.8 (6.4)	20–100 (47)
Confidence	74.2 (6.9)	20–100 (46)
Empathy	75.3 (6.3)	20–100 (35)
Test 3 MEARS, n = 600, median (IQR)	Median (IQR)	
Control	74 (70, 81)	20–126 (71)
Emotional non-defensiveness	82 (78, 103)	24–144 (70)
Faking	63 (59, 71.5)	20–126 (49)
Optimism	74 (70, 90)	20–126 (64)
Self-discipline	86 (82, 92)	20–126 (59)
Self esteem	77 (74, 84)	20–126 (57)
Test 4, n = 891	Mean (SD)	
Libertarian communitarian	78.3 (9.5)	30–120 (64)
Narcissism	28.6 (4.4)	10–50 (31)
Aloofness	22.6 (3.5)	10–50 (24)
Confidence	37.3 (3.7)	10–50 (23)
Empathy	36.9 (3.5)	10–50 (23)

the MEARS domains and EPM decile (r = 0.085–0.211). There was no significant correlation between any of the non-cognitive tests and the SJT score. Total UKFPO had weak correlation with the MEARS domains (r = 0.209 to 0.318). Of note, there was a strong correlation between student age and MEARS total (Spearman's r = 0.570. $p < 0.001$). (Not shown in tabular form).

As a large number of multi-variate analysis tests were performed, where significant results were obtained, the effects are quite small.

The multilevel analysis (see Table 4) shows that tests 1, 2 and 4 (libertarian-communitarian, NACE total and the abridged test 4) are not significantly associated with any of the four outcomes. END, part of MEARS is significantly associated with all four outcomes. Self-esteem is significantly associated with EPM decile and EPM total but the coefficients are very small. Aloofness and empathy domains in the NACE test are negatively associated with both SJT score and EPM decile.

In the MEARS domains, the emotional non-defensiveness (END: how one feels and reacts to people and situations) sub-test stood out as predicting all measures positively, with an accumulative effect such that a modest and achievable 7.5 extra marks (out of a valid range of 24–144) would improve total UKFPO score by 1 mark out of 100. Interestingly, increased self-esteem (out of 126) was related to a decrease in EPM decile and this filtered through to EPM total. One extra mark in aloofness (out of 50) led to a decrease in SJT score of 0.066 points, in other words, 15 extra aloofness marks led to a decrease in SJT of one point. Similarly, 14 extra points in empathy (out of 50) on average predicted one less SJT point.

Table 3 Univariate analysis: relationship between demographic variables and outcomes

Characteristics	Outcome variables							
	EPM decile	p value	EPM total	p value	SJT score	p value	UKFPO total	p value
Age at entry; rank correlation (r)	0.034	0.008	0.126	<0.001*	−0.054	<0.001*	−0.057	<0.001*
Gender								
Median (IQR)								
Male	5 (3, 8)	<0.001*	38 (41, 44)	<0.001*	40.4 (38.3, 42.3)	<0.001*	81.1 (76.9, 85.2)	<0.001*
Female	6 (4, 8)		38 (41, 44)		41.3 (39.1, 43.3)		82.4 (78.3, 86.4)	
Ethnic group								
Median (IQR)								
Caucasian	6 (4, 9)	<0.001*	41 (38, 44)	<0.001*	41.3 (39.2, 43.3)	<0.001*	82.8 (78.8, 86.6)	<0.001*
Non-Caucasian	5 (2, 7)		40 (37, 43)		39.9 (37.7, 41.9)		79.9 (75.8, 83.9)	
Type of secondary school attended								
Median (IQR)								
Fee-paying	6 (3, 8)	<0.001*	41 (38, 44)	0.872	41.1 (39, 43.2)	0.488	82.1 (78, 86.1)	0.776
Non-fee paying	6 (3, 8)		41 (38, 44)		41 (38.9, 43)		82.1 (77.7, 86.1)	
IMD quintile								
Median (IQR)								
Most affluent-I	6 (4, 8)	<0.001*	41 (38, 44)	<0.001*	41.3 (38, 44)	<0.001*	82.4 (78.4, 86.3)	<0.001*
II	6 (3, 8)		41 (38, 44)		41.3 (38, 44)		82.4 (78.3, 86.5)	
III	6 (3, 8)		41 (38, 44)		41.0 (38, 44)		81.8 (77.5, 85.6)	
IV	6 (3, 8)		41 (38, 43)		40.6 (38, 43)		81.0 (77.1, 85.3)	
Least affluent-V	5 (2, 8)		40 (38, 43)		40.2 (37, 43)		80.3 (76.5, 84.1)	
NS-SEC								
Median (IQR)								
NS-SEC categories								
1	6 (3, 8)	0.072	41 (38, 44)	0.006	41.1 (39.1, 43.1)	<0.001*	82.3 (78.4, 86.2)	<0.001*

Table 3 continued

Characteristics	Outcome variables							
	EPM decile	p value	EPM total	p value	SJT score	p value	UKFPO total	p value
2	6 (3, 8)		40 (37, 43)		41.1 (38.9, 42.9)		80.8 (78.3, 84.1)	
3	5 (3, 8)		41 (38, 44)		40.6 (38.3, 42.8)		81.5 (77.5, 84.9)	
4	5 (3, 8)		40 (37, 43)		39.9 (37.8, 42.2)		80.8 (77.1, 83.6)	
5	5 (3, 7)		39 (37, 42)		40.0 (37.8, 42.0)		78.7 (76.5, 83.6)	
Domicile								
Median (IQR)								
Domicile categories								
EU	6 (4, 8)	<0.001*	42 (40, 45)	<0.001*	40.2 (38.2, 42.5)	<0.001*	82.5 (79.2, 85.9)	<0.001*
International	4 (2, 7)		39 (36, 43)		39.3 (36.5, 41.3)		78.8 (73.9, 83.3)	
UK	6 (3, 8)		41 (38.8, 43.1)		41 (38.8, 43.1)		82.0 (77.9, 86)	
Total UCAS Score (r)	0.198	<0.001*	0.224		0.104		0.212	<0.001*

* Significant at the 5 % level following Bonferroni correction for multiple testing

Table 4 Multilevel analysis—non-cognitive test coefficients adjusted for year of UKCAT exam, gender, age at admission, school type attended and IMD quintile for the four outcomes

Independent variable	SJT score		EPM decile		Total EPM		UKFPO score	
	Estimate (95 % CI)	p value	Estimate (95 % CI)	p value	Estimate (95 % CI)	p value	Estimate (95 % CI)	p value
n = 721								
Libertarian-communitarian (TEST 1)	0.0055 (−0.015, 0.026)	0.591	0.0073 (−0.008, 0.023)	0.363	0.007 (−0.013, 0.027)	0.500	0.012 (−0.025, 0.05)	0.516
n = 774								
NACE total (TEST 2)	0.009 (−0.005, 0.023)	0.205	0.001 (−0.012, 0.014)	0.901	−0.001 (−0.016, 0.014)	0.914	0.009 (−0.011, 0.03)	0.370
n = 774								
Narcissism test	−0.022 (−0.058, 0.014)	0.238	−0.003 (−0.031, 0.025)	0.843	0.002 (−0.03, 0.034)	0.914	−0.019 (−0.073, 0.034)	0.475
Aloofness test	−0.066 (−0.122, −0.01)	0.022	−0.014 (−0.058, 0.03)	0.535	−0.013 (−0.062, 0.037)	0.621	−0.079 (−0.161, 0.004)	0.063
Confidence test	−0.002 (−0.042, 0.037)	0.913	−0.018 (−0.049, 0.013)	0.253	−0.013 (−0.048, 0.022)	0.456	−0.014 (−0.072, 0.045)	0.645
Empathy test	−0.071 (−0.116, −0.026)	0.002	−0.003 (−0.038, 0.032)	0.872	−0.003 (−0.043, 0.037)	0.879	−0.074 (−0.14, −0.007)	0.030
n = 472								
MEARS	0.013 (−0.017, 0.043)	0.409	−0.008 (−0.018, 0.003)	0.162	−0.005 (−0.017, 0.007)	0.427	−0.006 (−0.022, 0.01)	0.472
Control	−0.016 (−0.056, 0.024)	0.423	−0.019 (−0.056, 0.017)	0.295	−0.022 (−0.062, 0.019)	0.29	−0.031 (−0.093, 0.032)	0.337
Emotional non-defensiveness	0.055 (0.006, 0.104)	0.027	0.053 (0.009, 0.097)	0.019	0.074 (0.025, 0.123)	0.003	0.133 (0.056, 0.209)	0.001*
Faking	0.001 (−0.056, 0.059)	0.97	−0.041 (−0.093, 0.01)	0.117	−0.051 (−0.109, 0.006)	0.079	−0.05 (−0.14, 0.041)	0.281
Optimism	−0.03 (−0.077, 0.017)	0.211	−0.019 (−0.062, 0.024)	0.384	−0.035 (−0.083, 0.012)	0.142	−0.069 (−0.143, 0.006)	0.07
Self-discipline	−0.025 (−0.072, 0.021)	0.29	0.039 (−0.003, 0.081)	0.071	0.058 (0.011, 0.105)	0.016	0.025 (−0.048, 0.098)	0.507
Self esteem	0.002 (−0.05, 0.055)	0.926	−0.065 (−0.113, −0.017)	0.008	−0.076 (−0.129, −0.023)	0.005	−0.075 (−0.157, 0.008)	0.078
n = 708								
Libertarian-communitarian (TEST 4)	−0.022 (−0.054, 0.01)	0.172	−0.011 (−0.033, 0.011)	0.337	−0.006 (−0.031, 0.019)	0.641	−0.028 (−0.073, 0.016)	0.214
n = 705								

Table 4 continued

Independent variable	SJT score		EPM decile		Total EPM		UKFPO score	
	Estimate (95 % CI)	p value	Estimate (95 % CI)	p value	Estimate (95 % CI)	p value	Estimate (95 % CI)	p value
NACE total (Test 4)	0.022 (−0.013, 0.057)	0.202	0.011 (−0.010, 0.032)	0.316	0.006 (−0.018, 0.029)	0.653	0.025 (−0.016, 0.067)	0.231

* Significant at the 5 % level following Bonferroni correction for multiple testing

Discussion

This is the first study examining the predictive validity of paper and pencil tests of personality traits on admission to medical school against academic and non-academic outcomes on exit, in relation to both school-based and national performance indicators. We found some significant correlations but all with low effect sizes and an overall inconsistent picture. For example, aloofness and empathy scores on the NACE negatively predicted performance on the SJT but not the EPM decile or EPM total. Moreover, the actual patterns seem conflicting–higher empathy (representing emotional involvement) and higher aloofness (representing emotional detachment) both predicted performance in the same direction. Similarly, scores on the MEARS instrument generally lacked correlation although, first, it seemed the most sensitive test in that modest differences in scores could influence performance on the outcome measures, and, second, two scales appeared of interest. The emotional non-defensiveness (END: how one feels and reacts to people and situations) subtest stood out as predicting all outcomes measures positively while higher self-esteem was associated with lower EPM decile and EPM total scores. EPM is an indicator of academic achievement, mostly test performance, both written and clinical, but this does fit with the wider, non-medical literature which highlights that non-cognitive attributes can influence cognitive test performance (e.g., Gutman and Schoon 2013). However, these tests are not primarily being employed to predict academic performance and the small effect size with the SJT does not, on its own, seem sufficient to justify the use of such a test (although there may be an argument to explore the utility of the END sub-test further).

Where do our findings sit in comparison to previous literature? Powis and colleagues developed the Personal Qualities Assessment (PQA: which includes the IVQ and ITQ) and tested it in a number of centres. However, few of the reported studies have examined the predictive validity of the PQA, and those which have been carried out are limited in their methodology (e.g., small scale, local outcome measures e.g., Adams et al. 2012, 2015; Dowell et al. 2011; Manuel et al. 2005) and—at best—find only modest correlations (Adams et al. 2012, 2015). We would argue that, given the evidence to date as to the utility of SJTs in a variety of professional groups (see earlier, and Patterson et al. 2012a, b for a review) the use of a validated SJT as an outcome measure is more robust than the comparators used by other authors, and hence the weak relationship we found is probably a more accurate assessment of the power of the IVQ and ITQ to predict outcomes at the end of medical school.

It has been argued that the non-academic attributes the PQA measures are desirable in clinicians until the extremes are reached, as too much or little of any may be problematic. Indeed, Powis (2015) has gone as far as suggesting that the minority at these extremes might be excluded from the selection process. This view is not widely supported (e.g., Norman 2015) and indeed, given the low effect sizes and inconsistent picture we found with the NACE, we elected not to assess the 'extremes' as advocated by Powis and colleagues (e.g., Bore et al. 2009; Munro et al. 2008) as there seemed no justification for doing so. Certainly, on our evidence, the PQA cannot be justified as a tool or filter for excluding individual candidates.

Should we have expected there to be an association between performance on the various non-cognitive tests included in the UKCAT, and the EPM and the SJT? It could be argued that we compared apples and pears by expecting tests of personality traits to predict academic performance and the expression of job-specific personality traits in hypothetical situations. On the other hand, there is evidence that the "Big Five" personality factors correlate with academic performance at medical school (e.g., Lievens et al. 2002) and with

implicit trait policies (ITPs) (Motowidlo et al. 2006a, b). However, what about the additional influence of other personality traits such as motivation, resilience, perseverance, and social and communication skills (Gutman and Schoon 2013)? It was made clear to applicants that the non-cognitive tests within the UKCAT would not be used in selection decisions, so it would not be unreasonable to assume that those sitting this part of the UKCAT were less motivated to do well on these tests compared to the "high stakes" cognitive UKCAT tests. Conversely, the Foundation Programme application process is competitive so motivation to do one's best will be high.

There is also the issue of beliefs about the costs and benefits associated with expressing certain traits in particular situations. While ITP theory proposes to be related to individuals' inherent tendencies or traits, individuals must make judgements about how and when to express certain traits. Thus, SJTs are designed to draw on an applicant's knowledge of how they should respond in a given situation, rather than how they would respond. Although this seems a conceptual gap to us, there is some evidence that SJTs predict performance in one medical training context, that of UK general practice training (Lievens and Patterson 2011) (and the wider literature also suggests that the way an individual responds to an SJT question does predict actual behaviour and performance once in a role (e.g., McDaniel et al. 2001)). Validity studies have also shown that SJTs add incremental validity when used in combination with other predictors of job performance such as structured interviews, tests of IQ and personality questionnaires (O'Connell et al. 2007; McDaniel et al. 2007; Koczwara et al. 2012). While the focus of this paper is not to analyse the conceptual and theoretical frameworks of personality tools, it is essential that these are critically examined in order to develop, evaluate and compare medical selection tools and how these are used in admissions/selection processes.

This study is unusual in its scale, allowing for accurate estimates of correlations, subgroup analysis and multilevel modelling to more accurately estimate effect sizes. However, the range of outcome markers available was limited. The EPM is an indication of overall course academic achievement as judged against peers within each medical school: without a common exit exam it is not clear how much variation there is between schools and we are unable to estimate this effect, or correct for this. It is also a complex and varied measure as it includes other degrees and publications that will be confounded by age and other factors such as previous degrees. However, there are currently no comprehensive, standardized assessments across the UK akin to say the Canadian or US licensing examination, and so we had to be pragmatic and use what outcome measures were available to us. The SJT predictive validity remains to be determined but there is good reason to expect this, based on related previous work (McManus et al. 2013; Patterson et al. 2012a, 2015a). However, although we did not have access to the full dataset of test-takers (i.e. including those who either were not admitted to medical school or who did not graduate in 2013), mean scores and ranges across the non-cognitive tests were very similar between the full dataset summary (UKCAT technical reports 2007 and 2008) and the results from this graduating cohort (data not shown). In other words, those who graduated did not have significantly different non-cognitive scores from those who did, implying no range restriction due to subset selection. The non-cognitive tests were included in the 2007 and 2008 UKCAT test battery on a trial basis, and it was made clear that this data would not be used in decision making: this "low stakes" situation may have influenced candidate test behaviour, as discussed earlier (e.g., Abdelfattah 2010).

Norman (2015) argues that, without a clear negative relationship between academic achievement and desirable non-academic attributes, selection for medical school can and should seek students with attributes in both domains. To do so, requires valid, reliable and

affordable measurement techniques if we are to avoid an overly large initial filter on purely academic grounds. We must conclude that none of the non-cognitive tests evaluated in this study have been shown to have sufficient utility to be used in medical student selection in their current forms. Newer non-cognitive tests, such as the UKCAT entry level SJT (http://www.ukcat.ac.uk/about-the-test/situational-judgement/) will hopefully prove to be more useful in our context, when scrutinised in due course. We intend to follow up this cohort of doctors to examine the predictive validity of the cognitive and non-cognitive tests used at admission to medical school against post-graduate outcome measures.

Ethical permission

The Chair of the local ethics committee ruled that formal ethical approval was not required for this study given the fully anonymised data was held in safe haven and all students who sit UKCAT are informed that their data and results will be used in educational research. All students applying for the UKFPO also sign a statement confirming that their data may be used anonymously for research purposes.

Author contributions This study addressed a research question posed by a funding committee, of which JD was a member. JC and RMcK wrote the funding bid. JD advised on the nature of the non-cognitive data. RMcK managed the data and carried out the preliminary data analysis under the supervision of DA. DA advised on all the statistical analysis and carried out the multi-variate analysis. JC wrote the first draft of the introduction and methods sections of this paper. RMcK and DA wrote the first draft of the methods and results section, and JD the first draft of the discussion. JC and RMcK revised the paper following review by all authors.

Appendix: UKCAT non-cognitive test example questions

ITQ

For example:

Stems

1. I am aware of how frustrated I can get
2. I think others would describe me as easy going
3. I know I am more capable than most people
4. Others will talk, but I will act
5. I often feel dominated by others

Response options

A. definitely false
B. false on the whole
C. true on the whole
D. definitely true

IVQ

For example:

Situation

Peter and Jenny have known each other from childhood. Although from different families, they have always attended the same school and have lived next door to each other all their lives. They are as close as brother and sister. They are now in their final year of school.

In a mathematics exam, Peter happens to glance at Jenny who is sitting some three desks away and sees her take a sheet of paper from her coat pocket. Peter continues to stare and cannot believe what he is seeing – Jenny is cheating. Some time after the exam, a teacher approaches Peter and says "Jenny is in a lot of trouble. She has been accused of cheating, but I am certain she would not do that. You were sitting near her in the exam. Would you come with me to see the School Principal now and say that you saw no evidence of her cheating?"

What is your opinion? How do you feel about each of the following statements?

1. Close friends should always look after each other

    A strongly agree
    B agree
    C disagree
    D strongly disagree

2. Cheating is always wrong
    A strongly agree
    B agree
    C disagree
    D strongly disagree

3. It is important to get the best marks you can, whatever it takes
    A strongly agree
    B agree
    C disagree
    D strongly disagree

4. Some things are greater than friendships
    A strongly agree
    B agree
    C disagree

D strongly disagree

5. A good friend is always forgiving
   A strongly agree
   B agree
   C disagree
   D strongly disagree

6. The truth must always be told regardless of who might get hurt
   A strongly agree
   B agree
   C disagree
   D strongly disagree

MEARS

For Example:
   My behaviour is adapted to meet other's expectations.
   My behaviour is unaffected by other's expectations.
   Things usually turn out to be easier than I expected.
   Things usually turn out to be more difficult than I expected.
   Responses are on a six point scale from Strongly Agree to Strongly Disagree.

References

Abdelfattah, F. (2010). The relationship between motivation and achievement in low-stakes examinations. *Social Behavior and Personality, 2010*(38), 159–167.

Accreditation Council for Graduate Medical Education (ACGME). (2014). ACGME mission, vision and values. https://www.acgme.org/acgmeweb/tabid/121/About/Misson,VisionandValues.aspx.

Adams, J., Bore, M., Childs, R., Dunn, J., McKendree, J., Munro, D., et al. (2015). Predictors of professional behaviour and academic outcomes in a UK medical school: A longitudinal cohort study. *Medical Teacher, 2015*(37), 868–880.

Adams, J., Bore. M., McKendree, J., Munro, D., & Powis, D. (2012). Can personal attributes of medical students predict in-course examination success and professional behaviour? An exploratory prospective cohort study. *BMC Medical Education,* 12:69. http://www.biomedcentral.com/1472-6920/12/69/abstract.

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*, 313–321.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: a meta-analysis. *Personnel Psychology, 44*, 1–26.

Bore, M. R., Munro, D., Kerridge, I., & Powis, D. A. (2005a). Not moral "reasoning": A libertarian-communitarian dimension of moral orientation and Schwartz's value types. *Australian Journal of Psychology, 57*, 38–48.

Bore, M., Munro, D., Kerridge, I., & Powis, D. (2005b). Selection of medical students according to their moral orientation. *Medical Education, 39*, 266–275.

Bore, M., Munro, D., & Powis, D. (2009). A comprehensive model for the selection of medical students. *Medical Teacher, 31*, 1066–1072.

Childs R. (2012). Accessed 17th February 2015. http://www.teamfocus.co.uk/tests-and-questionnaires/understanding-motivation/resilience-scales.php.

Childs, R., Gosling, J., & Parkinson, M. (2008). Resilience Scales User's Guide Version 1. Accessed 18th February 2015. http://www.teamfocus.co.uk/user_files/file/Career%20Interests%20Inventory%20Users%20Guide%202013.pdf.

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgement tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology, 63*, 83–117.

Cleland, J. A., Dowell, J., McLachlan, J., Nicholson, S., & Patterson, F. (2012). Identifying best practice in the selection of medical students. http://www.gmc-uk.org/Identifying_best_practice_in_the_selection_of_medical_students.pdf_51119804.pdf.

Cleland, J. A., Patterson, F., Dowell, J., & Nicholson, S. (2014). How can greater consistency in selection between medical schools be encouraged? A mixed-methods programme of research that examines and develops the evidence base. http://www.medschools.ac.uk/SiteCollectionDocuments/Selecting-for-Excellence-research-Professor-Jen-Cleland-etal.pdf.

Dore, K. L., Kreuger, S., Ladhani, M., Rolfson, D., Kurtz, D., Kulasegaram, K., et al. (2010). The reliability and acceptability of the multiple mini-interview as a selection instrument for postgraduate admissions. *Academic Medicine, 85*, S60–S63.

Dowell, J., Lumsden, M. A., Powis, D., Munro, D., Bore, M., Makubate, B., & Kumwenda, B. (2011). Predictive validity of the personal attributes assessment for selection of medical students in Scotland. *Medical Teacher* 33, e485–e488 http://informahealthcare.com/doi/abs/10.3109/0142159X.2011.599448?prevSearch=allfield%253A%2528jon%2Bdowell%2529&searchHistoryKey.

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow trait. *Journal of Applied Psychology, 91*, 40–57.

Eva, K. W., Reiter, H. I., Rosenfeld, J., & Norman, G. R. (2004a). The ability of the multiple mini-interview to predict preclerkship performance in medical school. *Academic Medicine, 79*, S40–S42.

Eva, K. W., Reiter, H. I., Trinh, K., Wasi, P., Rosenfeld, J., & Norman, G. R. (2009). Predictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education, 43*, 767–775.

Eva, K. W., Rosenfeld, J., Reiter, H. I., & Norman, G. R. (2004b). An admissions OSCE: The multiple mini-interview. *Medical Education, 38*, 314–326.

Frank, J.R., & Snell, L. (2015). The draft CanMEDS 2015: physician competency framework. http://www.royalcollege.ca/portal/page/portal/rc/common/documents/canmeds/framework/framework_series_1_e.pdf.

Fukui, Y., Noda, S., Okada, M., Mihara, N., Kawakami, Y., Bore, M., et al. (2014). Trial use of the personal qualities assessment (PQA) in the entrance examination of a Japanese Medical University: Similarities to the results in western countries. *Teaching and Learning in Medicine, 26*, 357–363.

Gafni, N., Moshinsky, A., Eisenberg, O., Zeigler, D., & Ziv, A. (2012). Reliability estimates: Behavioural stations and questionnaires in medical school admissions. *Medical Education, 46*, 277–288.

Gale, T. C., Roberts, M. J., Sice, P. J., Langton, J. A., Patterson, F. C., Carr, A. S., et al. (2010). Predictive validity of a selection centre testing non-technical skills for recruitment to training in anaesthesia. *British Journal of Anaesthesia, 105*, 603–609.

General Medical Council. (2009). *Tomorrow's doctors: Outcomes and standards for undergraduate medical education*. GMC, London. http://www.gmc-uk.org/Tomorrow_s_Doctors_1214.pdf_48905759.pdf.

Girotti, J. A., Park, Y. S., & Tekian, A. (2015). Ensuring a fair and equitable selection of students to serve society's health care needs. *Medical Education, 49*, 84–92.

Gutman, L.M., & Schoon, I. (2013). *The impact of non-cognitive skills on outcomes for young people*. Institute of Education, London. https://educationendowmentfoundation.org.uk/uploads/pdf/Non-cognitive_skills_literature_review.pdf.

Hofmeister, M., Lockyer, J., & Crutcher, R. (2008). The acceptability of the multiple mini interview for resident selection. *Family Medicine, 40*, 734–740.

Hofmeister, M., Lockyer, J., & Crutcher, R. (2009). The multiple mini-interview for selection of international medical graduates into family medicine residency education. *Medical Education, 43*, 573–579.

James, D., Ferguson, E., Powis, D., Bore, M., Munro, D., Symonds, I., et al. (2013). Graduate entry to medicine: Widening psychological diversity. *BMC Med Education, 9*, 67.

Koczwara, A., Patterson, F., Zibarras, L., Kerrin, M., Irish, B., & Wilkinson, M. (2012). Evaluating cognitive ability, knowledge tests and situational judgement tests for postgraduate selection. *Medical Education, 46*, 399–408.

Lievens, F. (2013). Adjusting medical school admission: Assessing interpersonal skills using situational judgement tests. *Medical Education, 47*, 182–189.

Lievens, F., Coetsier, P., De Fruyt, F., & De Maeseneer, J. (2002). Medical students' personality characteristics and academic performance: A five factor model perspective. *Medical Education, 36*(11), 1050–1056.

Lievens, F., & Patterson, F. (2011). The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced level high-stakes selection. *Journal of Applied Psychology, 96*, 927–940.

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review, 37*, 426–441.

Lumsden, M. A., Bore, M., Millar, K., Jack, R., & Powis, D. (2005). Assessment of personal attributes in relation to admission to medical school. *Medical Education, 39*, 258–265.

Manuel, R. S., Borges, N. J., & Gerzina, H. A. (2005). Personality and clinical skills: Any correlation? *Academic Medicine, 80*, S30–S33.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgement tests, response instructions and validity: A meta-analysis. *Personnel Psychology, 60*, 63–91.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*, 730–740.

McManus, I. C., Woolf, K., Dacre, J., Paice, E., & Dewberry, C. (2013). The academic backbone: Longitudinal continuities in educational achievement from secondary school and medical school to MRCP(UK) and the specialist register in UK medical students and doctors. *BMC Medicine* 11(1), 242. http://www.biomedcentral.com/content/pdf/1741-7015-11-242.pdf.

Motowidlo, S. J., & Beier, M. E. (2010a). Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a situational judgment test. *Journal of Applied Psychology, 95*(2), 321–333.

Motowidlo, S. J., & Beier, M. E. (2010b). The effects of implicit trait policies and relevant job experience on scoring keys for a situational judgment test. *Journal of Applied Psychology, 95*, 321–333.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology, 75*(6), 640–647.

Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology, 91*(4), 749–761.

Munro, D., Bore, M. R., & Powis, D. A. (2005). Personality factors in professional ethical behaviour: Studies of empathy and narcissism'. *Australian Journal of Psychology, 57*, 49–60.

Munro, M., Bore, M., & Powis, D. (2008). Personality determinants of success in medical school and beyond: "steady, sane and nice". In S. Boag (Ed.), *Personality down under perspectives from Australia* (pp. 103–112). New York: Nova Science Publishers Inc.

Nedjat, S., Bore, M., Majdzadeh, R., Rashidian, A., Munro, D., Powis, D., et al. (2013). Comparing the cognitive, personality and moral characteristics of high school and graduate medical entrants to the Tehran University of Medical Sciences in Iran. *Medical Teacher, 35*, e1632–e1637.

Norman, G. (2015). Identifying the bad apples. *Advances in Health Sciences Education, 20*, 299–303.

O'Connell, M. S., Hartman, N. S., McDaniel, M. A., Grubb, W. L., & Lawrence, A. (2007). Incremental validity of situational judgement tests for task and contextual job performance. *International Journal of Selection and Assessment, 15*, 19–29.

O'Brien, A., Harvey, J., Shannon, M., Lewis, K., & Valencia, O. (2011). A comparison of multiple mini-interviews and structured interviews in a UK setting. *Medical Teacher, 33*, 397–402.

Patterson, F. (2013). Selection for medical education and training: Research, theory and practice. In K. Walsh (Ed.), *Oxford Textbook for Medical Education* (pp. 385–397). Oxford: Oxford University Press.

Patterson, F., Aitkenhead, A., Edwards, H., Flaxman, C., Shaw, R., & Rosselli, A. (2015a). Analysis of the situational judgement test for selection to the foundation programme 2015: Technical Report.

Patterson, F., & Ashworth, V. (2011). Situational judgement tests; the future for medical selection? *British Medical Journal.* http://careers.bmj.com/careers/advice/view-article.html?id=20005183.

Patterson, F., Ashworth, V., Mehra, S., & Falcon, H. (2012a). Could situational judgement tests be used for selection into dental foundation training? *British Dental Journal, 213*, 23–26.

Patterson, F., Ashworth, V., Zibarras, L., Coan, P., Kerrin, M., & O'Neil, P. (2012b). Evaluations of situational judgement tests to assess non-academic attributes in selection. *Medical Education, 46*, 850–868.

Patterson, F., Carr, V., Zibarras, L., Burr, B., Berkin, L., Plint, S., et al. (2009). New machine-marked tests for selection into core medical training: Evidence from two validation studies. *Clinical Medicine, 9*, 417–420.

Patterson, F. P., Kerrin, M., Edwards, H., Ashworth, V., & Baron, H. (2015) Validation of the F1 selection tools. Leeds: Health Education England, 2015. www.foundationprogramme.nhs.uk/download.asp?file=Validation_of_the_F1_selection_tools_report_FINAL_for_publication.pdf. Accessed 9 November 2015.

Patterson, F., Knight, A., Dowell, J., Nicholson, S., & Cleland, J. A. (2016). How effective are selection methods in medical education? A systematic review. *Medical Education, 50*(1), 36–60.

Powis, D. (2015). Selecting medical students. *Medical Teacher, 37*, 252–260.

Powis, D. A., Bore, M. R., Munro, D., & Lumsden, M. A. (2005). Development of the personal qualities assessment as a tool for selecting medical students. *Journal of Adult & Continuing Education, 11*, 3–14.

Prideaux, D., Roberts, C., Eva, K., Centeno, A., McCrorie, P., McManus, C., et al. (2011). Assessment for selection for the health care professions and specialty training: International consensus statement and recommendations. *Medical Teacher, 33*, 215–223.

Randall, R., Davies, H., Patterson, F., & Farrell, K. (2006a). Selecting doctors for postgraduate training in paediatrics using a competency based assessment centre. *Archives of Disease in Childhood, 91*, 444–448.

Randall, R., Stewart, P., Farrell, K., & Patterson, F. (2006b). Using an assessment centre to select doctors for postgraduate training in obstetrics and gynaecology. *The Obstetrician & Gynaecologist, 8*, 257–262.

Reiter, H. I., Eva, K. W., Rosenfeld, J., & Norman, G. R. (2007). Multiple mini-interviews predict clerkship and licensing examination performance. *Medical Education, 41*, 378–384.

Roberts, C., Walton, M., Rothnie, I., Crossley, J., Lyon, P., Kumar, K., et al. (2008). Factors affecting the utility of the multiple mini-interview in selecting candidates for graduate-entry medical school. *Medical Education, 42*, 396–404.

Rosenfeld, J. M., Reiter, H. I., Trinh, K., & Eva, K. W. (2008). A cost efficiency comparison between the multiple mini-interview and traditional admissions interviews. *Advances in Health Sciences Education, 13*, 43–58.

Rothmann, S., & Coetzer, E. P. (2003). The big five personality dimensions and job performance. *Journal of Industrial Psychology, 29*, 68–74.

Salgado, J. F. (1997). The five-factor model of personality and job performance in the European community. *Journal of Applied Psychology, 82*, 30–43.

Savin Baden, M., & Major, C. H. (2013). *Qualitative research: The essential guide to theory and practice.* Routledge, London. Cited in Cleland, J. A, & Durning, S. J. (2015) *Researching medical education.* Wiley, London.

Schuwirth, L., & Cantillon, P. (2005). The need for outcome measures in medical education. *British Medical Journal, 331*, 977.

ten Cate, O., & Smal, K. (2002). Educational assessment center techniques for entrance selection in medical school. *Academic Medicine, 77*, 737.

Ziv, A., Rubin, O., Moshinsky, A., Gafni, N., Kotler, M., Dagan, Y., et al. (2008). MOR: A simulation-based assessment centre for evaluating the personal and interpersonal attributes of medical school candidates. *Medical Education, 42*, 991–998.