



**How effective are selection methods in medical education?
A systematic review**

Journal:	<i>Medical Education</i>
Manuscript ID:	MED-2014-1177.R1
Manuscript Type:	Research Papers
Keywords:	Testing/Assessment, Admission/Selection/minority recruitment

SCHOLARONE™
Manuscripts

review

How effective are selection methods in medical education and training?

A systematic review

Introduction

It is essential to ensure that selection methods used by recruiters are robust as selection is the first assessment for entry into medical education and training, and medical school admissions internationally are highly competitive. There is also an ethical, political and economic responsibility for medical education and training to produce competent clinicians, due to the high-stakes nature of the profession with regards to individuals' and societies' health, well-being and financial cost. Krieter and Axelson's (1) non-systematic review of medical admissions research and practice in the last 25 years noted that effective educational interventions typically produce only small gains in learning (effect sizes generally below .20), whereas evidence-based selection is comparatively far more powerful, with well-designed selection tools achieving performance gains exceeding one standard deviation. Accordingly, a central concern is to determine which different selection methods can reliably identify those who will be successful in medical training and ultimately become competent clinicians.

Traditionally, selection for medicine has involved several different methods used in combination. Prior academic attainment is generally the primary basis for selection, which is usually assessed at an initial screening stage (2). Academic indicators are typically used as the basis for initial shortlisting decisions in combination with personal statements, references or aptitude tests or both, usually followed by an interview at the final stage to make selection decisions. However, there are several concerns about this approach. First, previous reviews have concluded that academic performance is a good, but not perfect, predictor of performance, accounting for approximately 23% of the variance in performance in undergraduate medical

1
2
3 training and 6% in postgraduate performance (3). It could be argued that academic ability is a
4
5 necessary but not sufficient condition to ensure that a trainee becomes a competent clinician.
6

7
8 Second, although academic achievement is consistently shown to be a good predictor of
9
10 performance in medical school (4), there has been substantially less attention paid to researching
11
12 methods that reliably evaluate important (non-academic) personal attributes, interests and
13
14 motivational qualities. It cannot be assumed that those with high academic ability alone can be
15
16 turned into competent physicians via medical training, as other skills and qualities may need to be
17
18 present from the start (5).
19

20
21 Third, there has been a dearth of longitudinal cohort studies examining the predictors of
22
23 success after qualification. Specifically, there is a research gap with respect to long-term follow-up
24
25 of trainees, linking performance on different selection methods with subsequent performance in
26
27 clinical practice.
28

29
30 Medical school admissions processes and selection for specialty training attract strong
31
32 public interest and often criticism regarding fairness (6-8). There is a pressing need to review the
33
34 research evidence of how best to design and validate selection methods and systems to guide
35
36 recruiters in future. Moreover, relatively little research has been conducted exploring the quality
37
38 and effectiveness of selection methods other than academic attainment, such as interviews,
39
40 personal statements and references.
41
42

43
44 In order to explore these issues, we report here the results of a new systematic search
45
46 and review of the research literature, examining studies in both undergraduate and postgraduate
47
48 medical education. Specifically, we present the existing data on the relative strength of the
49
50 research evidence underlying the quality of each of those methods as well as their findings to
51
52 shape a future research agenda and to inform future practice.
53
54

55 56 **Method** 57 58 59 60

Data Sources

We conducted a formalised literature search using the criteria specified in Table 1. Our results were limited to English-language studies published between January 1997 and August 2014.

INSERT TABLE 1 HERE

Study selection and inclusion and exclusion criteria

AK and FC reviewed the abstracts of all articles identified by the search to remove obviously irrelevant papers. Any articles that were potentially relevant were highlighted, and were reviewed for a second time by AK, FC and FP. AK and FP discussed these papers until both reviewers agreed about whether the paper should be included in the review. A standardised set of inclusion criteria was generated: papers should be peer-reviewed, and contain empirical data relating to selection into medical education or training. We also included relevant systematic and meta-analytic reviews and non-systematic critical reviews, but excluded general opinion pieces, commentaries and letters. After applying our inclusion criteria, duplicate papers were removed, leaving the remaining articles to be retrieved for full-text review. Three authors (AK, FC & FP) independently examined each of these articles for inclusion.

Assessment of study type, quality and selection method

Papers meeting the inclusion criteria were reviewed against three criteria: (1) selection method type (e.g. interview, selection centre, etc); (2) research question addressed (e.g. cost effectiveness, acceptability, etc - see Muir & Grey, 1996, cited in (9)); and (3) type of study design (e.g. meta-analyses, cross-sectional qualitative study). By assessing papers against these three criteria, we were able to make general statements about the quality of evidence available in relation to different research questions for different selection methods. To generate a list of the different selection methods, AK listed the selection method(s) assessed in each paper meeting the

1
2
3 inclusion criteria, and asked an independent researcher to check the papers against the list for
4
5 errors.

6
7
8 The research question and evidence quality categories are displayed in Table 2. In relation
9
10 to the different research questions under investigation, we removed Muir & Grey's (1996)
11
12 "salience" and "safety" categories, as they were not relevant to our context. We also combined the
13
14 "acceptability" and "appropriateness" categories, and refocused the "procedural issues" category
15
16 to more appropriately reflect the considerations given to implementing selection tools in medical
17
18 education. Therefore, we examined each study in relation to four research questions: *effectiveness*,
19
20 *procedural issues*, *acceptability* and *cost-effectiveness*. This approach was intended to address the
21
22 assumption implicit in much previous research that predictive validity is the most important
23
24 measure of the effectiveness of a selection method; as the authors acknowledge that the success of
25
26 a selection tool may be determined by a range of additional factors, including its accessibility, ease
27
28 of implementation and the extent to which it is viewed as acceptable by key stakeholders. Finally,
29
30 in relation to study quality, we categorised papers into five general study types, including
31
32 systematic and non-systematic reviews, longitudinal studies, and quantitative and qualitative
33
34 cross-sectional studies. Studies reported within meta-analyses and systematic reviews were not
35
36 assessed individually; rather these are reported as the overall findings of each meta-analysis or
37
38 systematic review.
39
40
41

42
43 ***INSERT TABLE 2 HERE***
44
45
46

47 48 **Results**

49
50 The literature search produced 1,407 hits across all databases including duplicates
51
52 (EBSCO = 732, Embase = 501, ERIC = 49, SCOPUS = 50, Web of Knowledge = 107). The titles
53
54 and abstracts of the 1,407 search results were scanned to remove obviously irrelevant articles and
55
56 duplicates (n=1,079), leaving 326 articles for review. These abstracts were screened according to
57
58
59
60

1
2
3 the eligibility criteria, removing a further 28 articles (see Figure 1). Two researchers (AK and FC)
4
5 made all decisions, but any uncertainties were discussed with another member of the research
6
7 team (FP). Copies of the 298 articles were obtained and examined. Review of the full text
8
9 removed a further 121 articles. A total of 179 articles met the inclusion criteria for the present
10
11 review¹.
12

13
14 ***FIGURE 1 FLOWCHART HERE***
15

16 The 179 studies were sorted into eight categories of different selection methods. Table 3
17
18 shows the number of papers returned in relation to each selection method (rows) and research
19
20 question (columns). Studies investigating multiple selection methods or research categories or
21
22 both were assigned to multiple categories, as required.
23

24
25 ***INSERT TABLE 3 HERE***
26

27 A summary of the relevant review findings are presented in Table 4. The authors acknowledge
28
29 that there is a range of the quality of studies presented, irrespective of the study type; however it is
30
31 beyond the scope of this review to provide a detailed account of the quality of each study.
32

33 Therefore Table 4 is intended to provide a brief overview of the research evidence, rather than to
34
35 provide a comprehensive description of each study.
36
37

38
39 ***INSERT TABLE 4 HERE***
40

41 We provide a more detailed overview of our synthesis of the research evidence below.
42

43 **1. Aptitude Tests**

44
45 (a) *Type of evidence.* Fifty studies were reviewed. Of these, three were systematic reviews/ meta-
46
47 analyses, three were non-systematic reviews, 31 were longitudinal (one was a meta-analysis), and
48
49 thirteen were cross-sectional (one mixed method, one tool development, one qualitative, and ten
50
51 quantitative).
52

53
54
55
56
57

¹ The results section provides a summary of the evidence from the literature. For a full list and description of all
58
59 papers identified in the review, refer to Tables 3 and 4.
60

1
2
3 (b) *Effectiveness*. There is mixed evidence on the predictive validity of aptitude tests in medical
4 student selection. Some researchers have presented evidence to support the reliability and
5 criterion, incremental or predictive validity for aptitude tests including the MCAT (10-13),
6 GAMSAT (14), UMAT (15, 16), HPAT (4), UKCAT (17-20), BMAT (21, 22), Qudraat (23), and
7 a surgical aptitude test for practical skills for admission to a Otolaryngology residency programme
8 in the USA (24). Other researchers are sceptical of the reliability or effectiveness of the MCAT
9 (25), UKCAT (26), GAMSAT (27), UMAT (28-33), BMAT (34, 35), and an unspecified aptitude
10 test (36). However, some evidence suggests that students selected using an aptitude test may be
11 more able and better motivated to study medicine than those selected using a process not including
12 an aptitude test (37). Finally, one paper (34) reported a nuanced finding that section one (science
13 knowledge and applications) of the BMAT was predictive of medical school performance, while
14 section two (aptitude and skills) was not.

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30 (c) *Procedural issues*. Research suggests that variations in the way that aptitude tests are used in
31 medical student selection may affect their reliability or validity (38-41). This is notable as medical
32 schools vary in how they use aptitude tests to inform selection decisions, and the statistical
33 methods they use for determining cut-scores and predicting subsequent performance. One article
34 (40) reported that the dimensionality of an aptitude test affected its effectiveness as a selection
35 tool, with a scale composed of three subject-specific dimensions (biology, physics and chemistry)
36 having better psychometric properties than a uni-dimensional model, even with the subject-
37 specific scales being highly correlated and being used to calculate a global score.

38
39
40
41
42
43
44
45
46
47 (d) *Acceptability*. One study (12) reported that aptitude test scores were one of the most influential
48 factors determining decisions made by medical school admissions committee members. However,
49 another study (42) reported that few first year medical students agreed that the aptitude tests were
50 a useful part of the selection procedure to medical school.
51
52
53
54
55
56
57
58
59
60

1
2
3 (e) *Cost effectiveness*. No papers were reviewed that address the cost effectiveness of aptitude
4 tests.
5

6
7 (f) *Summary*. Mixed evidence exists among researchers on the usefulness of aptitude tests in
8 medical student selection which largely depends on the specific aptitude test studied, such that
9 generality of findings is problematic. For example, some studies support the predictive validity of
10 aptitude tests and other research suggests the selection method lacks predictive validity. In
11 particular, the UMAT has been subject to significant criticism in this regard in recent years. Mixed
12 evidence also exists on the fairness of aptitude tests, with some research suggesting that certain
13 groups score more highly on aptitude tests than other groups, while other research suggests that
14 this is not the case. For example, there is mixed evidence on the equity of aptitude tests for
15 different groups of medical school applicants (e.g. sex, age, language status, and socio-economic
16 status) (10, 18, 22, 43-47). Other evidence suggests that aptitude tests are equitable with respect to
17 candidate background, are affected relatively little by candidate coaching, and remain stable over
18 time (18, 22, 47-49), with the possible exception of the UMAT (29). It is therefore important to
19 evaluate each aptitude test in their own right in order to draw conclusions regarding the quality of
20 the tool.
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 **2. Academic Records**

41
42 (a) *Type of evidence*. Thirty-one studies were identified which assessed academic records. Twenty
43 four of these were longitudinal (one was a meta-analysis), two were meta-analyses, one was a non-
44 systematic review, and four were cross-sectional, quantitative.
45
46
47
48

49 (b) *Effectiveness*. Research evidence is generally highly concordant and supports the predictive
50 validity of academic records in medical student selection (6, 14, 15, 23, 27, 32, 50-56). McManus
51 and colleagues (57) describe how prior educational attainment forms the academic backbone of
52 selection, progression through medical school and beyond. Another paper describes a small but
53
54
55
56
57
58
59
60

1
2
3 significant incremental validity gain through using candidates' educational achievement alongside
4 aptitude tests compared to the use of traditional academic indicators alone (18). International
5 evidence also suggests that candidates admitted on the basis of their academic record had lower
6 levels of dropout than those who were not (58, 59). Incremental validity may be provided through
7 the addition of an appropriate aptitude test (3, 19, 60). A minority of studies (17, 36, 61) reported
8 that academic records were not predictive of medical school performance.

9
10
11
12
13
14
15
16 (c) *Procedural issues*. Some authors have argued that academic records may be unstable or
17 lacking in sufficient power for making fine distinctions between candidates (49, 62, 63). For
18 example, McManus and colleagues (62, 63) posited that the current grading system of A-Levels in
19 the UK does not offer sufficient discriminatory power to enable the selection of the most able
20 students.

21
22
23 (d) *Acceptability*. Evidence was mixed on the acceptability of using academic records in medical
24 student selection. This is illustrated by some authors citing academic records as an important
25 factor that can influence selection decisions (12).

26
27
28 (e) *Cost effectiveness*. No papers were reviewed that address the cost effectiveness of academic
29 records.

30
31
32
33
34 (f) *Summary*. A high level of consensus exists among researchers that academic records provide
35 useful information to inform medical student selection. Research generally suggests prior
36 academic attainment has predictive power, meaning those with stronger academic records are
37 more likely to succeed in medical school. However, there is concern that the discriminatory power
38 of prior academic attainment may be diminishing as increasing numbers of medical school
39 applicants have top grades. There is also a lack of long-term follow-up data to provide evidence
40 that medical school applicants with higher grades go on to become better doctors. Moreover,
41 Milburn (2012) notes that over-reliance on A-Level results may create a distorted social intake to
42 universities, and recruiting medical students solely on the basis of academic attainment may
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 neglect important non-cognitive factors required for success in medical school and beyond.
4
5 Further research is required to gauge the extent to which this is an international problem.
6
7
8

9 10 **3. Personal Statements**

11 (a) *Type of evidence.* A total of fifteen studies were reviewed, four of which were longitudinal.
12
13 The remaining studies were cross-sectional (three qualitative, seven quantitative), and one was a
14
15 non-systematic review.
16

17
18 (b) *Effectiveness.* Evidence is mixed on the predictive validity of personal statements. Although
19
20 some evidence has been found for the predictive validity of personal statements for medical school
21
22 drop-out rates (59), performance on internal medicine (13), and clinical aspects of training (60),
23
24 others have reported that personal statements have low reliability compared to other common
25
26 selection instruments (64) and were not predictive of subsequent success at a medical school (1,
27
28 65, 66)
29

30
31 (c) *Procedural issues.* Evidence suggests that a number of procedural factors affected the
32
33 reliability and validity of personal statements. Medical school candidates may use personal
34
35 statements to present themselves in ways they believe are attractive to admission committees,
36
37 which may not necessarily be accurate (67, 68). As such, the information captured by personal
38
39 statements is likely to be both partial and subjective in nature. Factors that may affect the
40
41 effectiveness of the selection method include earliness of submission in relation to a deadline (69),
42
43 marking method, and on-site versus off-site completion (70). Finally, one article highlighted the
44
45 fact that personal statements are used differentially by different UK medical schools (71): some
46
47 medical schools formally used the information in making selection decisions, while others ignore
48
49 this information due to concerns that it may unfairly bias selection decisions.
50
51

52
53 (d) *Acceptability.* Research has highlighted potential sources of data contamination in personal
54
55 statements, including candidates' prior expectations, length of time spent completing submissions,
56
57
58
59
60

1
2
3 and input in submissions from third parties. Other research (12, 67) has commented on the
4 political validity and stakeholder satisfaction of personal statements in medical student selection.
5
6
7 Elam et al. (12) reported that the contents of medical school candidates' application forms are very
8 unlikely to exert any significant influence on decisions made by admissions committees. White
9 and colleagues (67) also argued that medical school candidates present themselves in ways that
10 they believe are expected of candidates, rather than in ways that are a genuine reflection of
11 themselves. Likewise, Kumwenda and colleagues (72) found that most medical school applicants
12 believed that others stretched the truth in their personal statements, and a proportion of applicants
13 believed that it is unlikely that they were checked for accuracy.

14
15
16
17
18
19
20
21
22
23 (e) *Cost effectiveness*. No papers were reviewed that address the cost effectiveness of personal
24 statements.

25
26
27
28 (f) *Summary*. Evidence on the effectiveness of personal statements in medical student selection is
29 mixed at best. Some evidence exists to support the predictive validity of personal statements.
30 However, a large volume of research evidence suggests that the selection method lacks reliability
31 and validity. Personal statements remain widely used in medical school selection worldwide,
32 despite concerns that the effectiveness of the selection method is influenced by numerous
33 extraneous factors. The content of personal statements may also unfairly cloud the judgement of
34 individuals making selection decisions.
35
36
37
38
39
40
41
42
43
44

45 **4. References**

46
47 (a) *Type of evidence*. A total of six articles were reviewed: two were non-systematic reviews, two
48 were longitudinal, one was qualitative cross-sectional and the last was quantitative cross-sectional.

49
50
51 (b) *Effectiveness*. Studies examining the effectiveness of references did not usually include a
52 direct empirical test of predictive validity (12, 60, 71, 73, 74), although there was some direct
53 evidence (60) that this selection method did not consistently predict performance at medical
54
55
56
57
58
59
60

1
2
3 school. Nevertheless, there was clear consensus among researchers that referees' reports were of
4
5 limited use in predicting performance at medical school.
6

7 (c) *Procedural issues*. One study (75) examined referees' reports, and found that the content of the
8
9 reports made it impossible for admissions committees to differentiate between applicants on the
10
11 basis of the data they contain. Therefore, the authors concluded that the utility of referees' reports
12
13 in medical student selection is questionable at best.
14

15 (d) *Acceptability*. Direct assessments of the acceptability of references were critical of the
16
17 inclusion of referees' reports in medical student selection, and remarked that the information they
18
19 contain may unduly bias admissions committees. For example, Ferguson et al. (60) found that the
20
21 information in teachers' references did not consistently predict medical school performance, and
22
23 Poole et al. (74) claimed that personal references have no predictive value. One study commented
24
25 that referees' reports remain widespread in medical student selection (71).
26
27

28 (e) *Cost effectiveness*. No papers were reviewed that address the cost effectiveness of references.
29

30 (f) *Summary*. There is a good level of consensus that references are neither a reliable nor valid
31
32 method for selecting candidates applying for medical school. Despite these findings, references
33
34 remain a common feature of medical school selection worldwide. To this extent, the inclusion of
35
36 references in medical school admissions processes may be unhelpful and use valuable resources
37
38 that could be directed more usefully to selection methods with evidentially based reliability and
39
40 validity.
41
42
43
44
45

46 47 **5. Situational Judgement Tests**

48 (a) *Type of evidence*. A total of 24 studies were reviewed. Of these, eight were longitudinal, five
49
50 were cross-sectional quantitative studies, four were systematic reviews, and five were non-
51
52 systematic reviews. Of the remaining two studies, one was developing a psychometric test, and the
53
54 other was a multiple cohort study.
55
56
57
58
59
60

1
2
3 (b) *Effectiveness*. Despite some concern about their susceptibility to coaching (76), overall there is
4 a good level of consensus among researchers that situational judgement tests (SJTs) are a reliable
5 and valid selection method across a range of occupations, including selection of medical students
6 (77-85).
7
8

9
10
11 (c) *Procedural issues*. Research suggests that the mode of administration may affect SJTs, with
12 video-based SJTs having higher operational validities than equivalent paper-and-pencil SJTs (58,
13 5) Similarly, different response instructions and methods of constructing alternative forms may
14 affect the validity of the SJT selection method (86, 87). In terms of equity, mixed evidence exists
15 on the relative susceptibility of SJTs to coaching (82, 88).
16
17
18
19
20
21

22 (d) *Acceptability of SJTs*. Across four studies, SJTs were rated favourably as selection tools by
23 candidates (82, 89-91). Some evidence has been presented that mode of administration may affect
24 candidate evaluations of SJTs, with video-based SJTs rated more favourably than paper-and-
25 pencil SJTs (89). No studies were identified that examined the political validity or stakeholder
26 acceptance of SJTs in medical student selection.
27
28
29
30
31
32
33

34 Six studies were identified examining the appropriateness of SJTs as a component of a
35 wider selection process (82, 83, 92-95). The weight of evidence across these studies suggests that
36 SJTs can usefully be incorporated into selection procedures across numerous occupational groups.
37
38

39 (e) *Cost effectiveness*. One study (82) concluded that there was tentative evidence of the relative
40 cost-effectiveness of SJTs compared with other methods of assessment, although direct evidence
41 in this area was not presented. Cost is also an important consideration when comparing text-based
42 and video-based SJTs, given that video-based SJTs require significantly greater time and financial
43 resources to develop.
44
45
46
47
48
49
50

51 (f) *Summary*. There is a good level of consensus among researchers that SJTs, when properly
52 constructed, can form a reliable, valid, cost effective and acceptable element of medical school
53 selection systems. SJTs are a complex selection instrument, with a wide range of options available
54
55
56
57
58
59
60

1
2
3 in relation to item formats, instructions and scoring. When these options are calibrated
4
5 appropriately, research evidence points to the strength of SJTs in medical student selection for
6
7 assessing non-cognitive attributes.
8
9

10 11 **6. Personality and Emotional Intelligence**

12
13
14 (a) *Type of evidence.* In total, 20 studies assessed personality and six assessed emotional
15
16 intelligence. Of the personality studies, eight were longitudinal (one was a meta-analysis), five
17
18 were non-systematic reviews and seven were cross-sectional, quantitative. Two emotional
19
20 intelligence studies were longitudinal, one was a systematic review, and the other three were
21
22 cross-sectional, quantitative.
23

24
25 (b) *Effectiveness.* Despite some research finding no evidence for associations between personality
26
27 traits and medical school performance (96), a number of studies have found that the Big Five
28
29 personality traits (openness, conscientiousness, extraversion, agreeableness and neuroticism) may
30
31 correlate with various aspects of medical school performance (97). Conscientiousness, for
32
33 example, has also been shown to be a positive predictor of pre-clinical knowledge and exam
34
35 results (56, 60, 65, 98) and to offer incremental validity over knowledge-based assessments (60,
36
37 65). However, conscientiousness has also been found to be a significant negative predictor of
38
39 clinical performance (56, 65) demonstrating that the association between personality traits and
40
41 performance in medical education and training is complex and possibly non-linear. Indeed,
42
43 Ferguson et al (56) suggest that while personality research has long suggested that
44
45 conscientiousness is beneficial when selecting into organisations, it has a 'dark side', where for
46
47 example the facets of being methodical and dutiful may hinder the acquisition of knowledge in the
48
49 clinical years of medical school. "Dysfunctional" personality traits in medical students (including
50
51 paranoid, avoidant, passive aggressive, antisocial, narcissistic and uncooperative) have been
52
53 reported to be associated with lower academic grades (99, 100). Considering personality
54
55
56
57
58
59
60

1
2
3 assessment more broadly, it has also been demonstrated to provide incremental validity over
4
5 cognitive methods in a medical school selection process (101).
6

7
8 Two studies (102, 103) provide tentative evidence that Emotional Intelligence (EI) may be
9
10 an important ability for medical students that is usually not assessed by typical medical school
11
12 selection methods (104). Other studies found no significant correlations between EI and skill in
13
14 medical students (105, 106) or other selection procedures for medical school admission (107).
15
16 There is provisional evidence that a self-report measure of EI (WLEIS) does not significantly
17
18 correlate with measures of success in medical school, but an ability-based measure of EI
19
20 (MSCEIT) does (108). However, Cherry and colleagues (109) conclude that there is currently
21
22 insufficient evidence to support the use of EI as a selection method.
23

24
25 (c) *Procedural issues*. Lievens and colleagues (110) suggested that the validity of personality
26
27 measures in predicting medical school grades increases over the course of medical education and
28
29 training. Their finding that conscientiousness is an increasing asset for medical students as their
30
31 course becomes more clinical is in direct contrast to the findings reported by Ferguson and
32
33 colleagues (56, 60). This difference may be due to different populations and study designs, but it
34
35 may be that previous studies relying on early outcome criteria might have underestimated the
36
37 predictive value of personality variables. Although there are concerns that personality tests may be
38
39 ‘fakeable’, Hojat and colleagues (97) argue that their operational validity may be maintained by
40
41 reminding respondents to reply truthfully and that intentionally false responses can be detected by
42
43 a social desirability scale.
44
45

46
47 (d) *Acceptability*. Evidence is mixed as to the acceptability of personality assessment in medical
48
49 student selection (110). While positive evidence on the predictive validity of personality
50
51 assessment suggests that it is an appropriate and acceptable method for selecting medical students,
52
53 others (111) have cautioned against the adoption of personality measures without consideration of
54
55
56
57
58
59
60

1
2
3 potential future impacts on diversity in medical student personalities. No evidence was found on
4
5 the acceptability of EI in medical student selection.
6

7 (e) *Cost effectiveness*. Knights and Kennedy (100) concluded that measures of dysfunctional
8
9 personality types could usefully and cost-effectively be incorporated into medical student
10
11 selection. Similarly, Powis and Rolfe (112) gave consideration to the costs and benefits of the
12
13 selection procedure at a single medical school, but did not provide any direct evidence on the cost-
14
15 effectiveness of personality measures in medical student selection. No evidence was found on the
16
17 cost effectiveness of EI in medical student selection.
18
19

20 (f) *Summary*. Taken broadly, there is a relatively high level of consensus among researchers that
21
22 some domains or traits of personality are significantly positively or negatively associated with
23
24 aspects of performance in medical school. However, the associations between personality domains
25
26 and medical school performance are often complex, demonstrated by evidence that
27
28 conscientiousness may be positively associated with knowledge-based assessment, but negatively
29
30 associated with some clinical aspects of medical school assessment. This suggests that closer
31
32 attention to the criterion constructs should also be considered when reviewing personality-based
33
34 selection tools. Personality assessment can be cost-effective and best used in combination with an
35
36 interview method where applicant responses can be probed further. Recruiters should be aware
37
38 that there is a relative dearth of evidence regarding the long-term predictive validity of personality
39
40 assessment beyond medical school, and that there has been some concern that personality
41
42 assessment may narrow the diversity of types of individuals entering medical education and
43
44 training. Research on the predictive validity of EI assessment was sparse and at a very early stage
45
46 of development. The studies and reports were typically pilot studies or opinion pieces citing
47
48 evidence as to why EI may represent a valuable tool in future medical student selection processes.
49
50
51
52
53
54

55 56 **7. Interviews and Multiple Mini Interviews** 57 58 59 60

1
2
3 (a) *Type of evidence.* Seventy studies were found which assessed interviews. Of these, twenty-one
4
5 were longitudinal, one was a systematic review and four were non-systematic reviews. The
6
7 remaining studies were cross-sectional: four qualitative, one mixed-methods and 39 quantitative.
8

9
10 (b) *Effectiveness.* Despite some evidence to the contrary (13, 14, 32, 113-120) the balance of
11
12 evidence suggests that traditional interviews are generally not robust methods for selecting
13
14 medical students, and lack predictive validity (3, 8, 27, 73, 121-127) with Edwards and colleagues
15
16 (15) finding that poorer interview performance was associated with greater medical school GPA.
17
18 The mixed findings on the effectiveness of interviews may reflect the broad range of traditional
19
20 interview methods, from relatively unstructured individual interviews, to highly structured panel
21
22 interviews. However, Eva and Macala (128) found no difference in the reliability of interviewer
23
24 ratings between unstructured and structured multiple mini interview (MMI) stations, although
25
26 behavioural indicator stations differentiated between candidates more reliably than other station
27
28 types.
29
30

31
32 The findings from research on MMIs tend to be more directionally consistent than
33
34 research on traditional interviews: for example, the psychometric properties of MMIs are usually
35
36 reported to be adequate (129-134). However, Hissbach and colleagues (135) found that rater bias
37
38 had a greater effect on applicant scores than systematic differences in candidate performance.
39
40 There is little clarity about what is being measured within the different approaches described and
41
42 tightly standardised face-to-face interviews may not be comparable with scenario based MMI
43
44 stations utilising standardised role actors.
45
46

47
48 Consistent evidence is emerging of the predictive validity of MMIs, when exploring the
49
50 correlation between performance on MMIs and subsequent performance on both undergraduate
51
52 and postgraduate Objective Structured Clinical Examinations (133, 136-140) and other
53
54 examinations (66, 141, 142).
55
56
57
58
59
60

1
2
3 (c) *Procedural issues*. Schools differ significantly in terms of the length, panel composition,
4 structure, content and scoring methods of interviews. The differential usage of the interview
5 method in medical student selection may underlie the mixed findings regarding both reliability and
6 validity of interviews as reported above. Other research evidence suggests that candidate
7 performance may be significantly affected by coaching (29). Using interviews in a selection
8 process also presents logistical difficulties relating to the range and type of questions (143) and
9 interviewer subjectivity (48, 133, 144, 145).
10
11

12 (d) *Acceptability*. Most research reports that applicants and interviewers tend to view the
13 interviewing process positively, with tentative evidence that MMIs and more structured interviews
14 are preferred over less structured methods (128, 146). Some evidence exists suggesting that
15 aspiring medical students may prefer the schools that conduct interviews (147). Campagna-
16 Vaillancourt and colleagues (134) found that the majority of applicants and assessors perceived an
17 MMI appropriate to assess a range of competencies and was a fair process, as well as being
18 preferred over a traditional interview. Staged introduction of an MMI into a selection process may
19 foster institutional acceptance of the method (148). Standardised interviews can also be adapted
20 for use in postgraduate medical selection to measure characteristics that are considered important
21 and acceptable both to international medical graduates and interviewers (129, 131, 149).
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 (e) *Cost effectiveness*. The cost effectiveness of MMIs is generally reported to be good (142)
41 although comparatively interviews are significantly more costly than machine-marked tests.
42 Value-for-money may be further improved by examining the number of stations in an MMI, and
43 reducing the number of stations if reliability is not affected. However, some research suggests that
44 increasing the number of questions in MMIs increases reliability (133, 150). Indeed, the authors
45 estimated that to reach a Cronbach's coefficient alpha of .80 for high stakes assessment, MMIs
46 need 14 stations manned by a single interviewer. This number could be reduced to between seven
47 and 12 stations, if manned by two interviewers. Alternatively, Dodson and colleagues (151) found
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 that reducing the duration of MMI stations from eight to five minutes conserves resources with
4
5 minimal effect on applicant ranking and test reliability.
6

7 Tiller and colleagues (152) found that cost and time savings for candidates were
8
9 substantial when conducting an MMI online via Skype rather than in person, although further
10
11 research is required regarding the impact on fidelity, in terms of not having a face-to-face
12
13 encounter.
14

15
16 (f) *Summary.* Interviews are among the most widely used selection method for medical school
17
18 admissions. Evidence suggests that traditional interviews lack the reliability and validity that
19
20 would be expected of a selection instrument in a high stakes selection setting. Evidence also
21
22 suggests that the MMI offers improved reliability and validity over traditional interview
23
24 approaches. Further study is warranted in relation to the reliability of the MMI method, and its
25
26 predictive validity, particularly with respect to which attributes can be assessed reliably (e.g.
27
28 communication, critical thinking, empathy, etc.). More evidence is required as to the
29
30 appropriateness of criteria that can be assessed in interviews, informed by validation studies. In
31
32 addition, the cost efficiency and utility of MMIs should be evaluated, along with alternative
33
34 approaches to scoring and alternative uses of scores (including any minimum threshold criteria).
35
36 The use of MMIs has spread rapidly in recent years as they can be designed to be a reliable
37
38 selection method. However, issues surrounding the construct validity of MMIs remain
39
40 problematic: it is critically important that schools better understand what they are seeking to
41
42 measure, and actually are measuring, with this approach. The impact of MMI on candidates (in
43
44 terms of fairness, performance, coaching effects, etc.) is an outstanding practical concern that
45
46 should influence design decisions such as question rotation.
47
48
49
50

51 52 53 54 **8. Selection Centres** 55 56 57 58 59 60

1
2
3 (a) *Type of evidence.* A total of seven studies assessed selection centres (SCs). One of these was
4
5 longitudinal, and six were cross-sectional, quantitative.
6

7 (b) *Effectiveness.* Provisional evidence has been presented that SC methods may be reliable and
8
9 internally valid for assessing applicants' aptitude for medicine (153-155) and have predictive
10
11 validity for performance in postgraduate speciality training (156-158).
12

13 (c) *Procedural issues.* Implementing an SC as part of a process for selecting medical students may
14
15 be logistically complex. It requires the recruitment and training of faculty raters, and on-going
16
17 collaboration among academic and professional institutions and experts in different operational
18
19 aspects of the process (including simulation, evaluation and measurement) (155, 159). Moreover,
20
21 as SCs are based on a multi-trait, multi-method design, SCs may comprise a large number of
22
23 elements in different combinations and orders, meaning that the processes by which an SC is
24
25 designed and administered may influence the utility of the method.
26
27

28 (d) *Acceptability.* Provisional evidence exists that an SC for entry into specialty training was rated
29
30 favourably by candidates and assessors (156-158).
31
32

33 (e) *Cost effectiveness.* Evidence is mixed on the cost effectiveness of the SC method. It could be
34
35 argued that SCs can offer a cost-effective method of high-volume assessment for selection into
36
37 medical specialty training when balanced against the increased validity (and thus reduced
38
39 extended training costs) that SCs might offer. Ziv and colleagues (155) have shown that the SC
40
41 method can be expensive compared to other selection methods (approximately 300 USD per
42
43 candidate) and represents a logistically complex option, although on balance they still advocate
44
45 SCs for use in medical school selection. Roberts and colleagues (159) investigated the feasibility
46
47 of using healthcare staff participating in the simulation scenarios as raters, to minimise the human
48
49 resource required to implement an SC. However, staff participant ratings were different from those
50
51 of trained assessors, and failed to achieve adequate levels of inter-rater reliability. Nonetheless,
52
53
54
55
56
57
58
59
60

1
2
3 Roberts et al concluded that it may be viable to use other healthcare staff rather than trained
4 assessors for some but not all stations.
5

6
7 (f) *Summary*. Overall, research on the utility of SCs for medical student selection was relatively
8 sparse. Evidence on predictive validity for postgraduate selection is stronger although further
9 evidence is required to build a case for their predictive validity in medical school selection.
10
11
12

13
14
15
16 In Table 5, we summarise our review regarding the “evidential weight” and relevance
17 for each of the selection methods reviewed.
18

19
20
21 ***INSERT TABLE 5 HERE***
22

23 Discussion

24 Summary of Key Findings

25
26
27 Our review of a very broad literature identifies that research into medical selection
28 represents, to some extent, a picture of quantity over quality: a substantial number of studies are of
29 moderate quality at best with also some significant gaps in the reporting and evaluation of some
30 selection techniques. There is an over-reliance on cross-sectional study designs and a general
31 focus on reliability estimates as indicators of quality rather than aspects of validity (a method may
32 have high reliability but be “reliably wrong” (23)). Although there are some studies addressing
33 issues relating to predictive validity, there is very little research exploring construct validity issues
34 (i.e. what is being measured?) and the relative cost effectiveness of selection methods. Similarly,
35 there are very few long-term evaluation studies, and few examining the relative contribution of
36 various selection methodologies (and the impact of various weightings) when methods are used in
37 combination (as is the norm in medical school selection) (160, 161). It is hard to see how
38 substantial progress can be made without appropriately conceived and long-term studies to
39 systematically assess potentially promising approaches. This paper has sought to identify specific
40 areas where such work should be prioritised.
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 There are however, some clear messages about the comparative reliability, validity and
4 effectiveness of various methods. The academic attainment of candidates remains a common
5 feature of most selection policies and the strength of evidence for continuing to do so remains
6 strong. The extant evidence paints a relatively clear picture regarding structured interviews/MMIs,
7 SJs and SCs being more effective across several criteria and generally fairer than traditional
8 interviews, references and personal statements. Evidence is currently mixed regarding the
9 effectiveness and fairness of aptitude tests depending on the tool in question. Similarly, more
10 long-term validity evidence is required in exploring personality assessments. The picture
11 regarding the acceptability of various selection methods is also mixed, and may be influenced by a
12 variety of factors, including differing stakeholder views, variation in the philosophies of both
13 medical students and medical schools, and the way that the tool is implemented as part of a
14 selection system. This area would benefit from further exploration of the reasons driving the
15 acceptability of different selection methods.
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

32 When judging the papers in this review, it was clear that some terms cover a broad spectrum of
33 methods: MMIs, SJs, aptitude tests, personality assessments and SCs are measurement methods
34 and within each category comprise a multitude of different design parameters. For example, there
35 are many different types of interviews, even when structured. Even when considering MMIs,
36 personality test and SJs, the construction and content of the interview or test can vary
37 significantly. Depending on the design, this may significantly alter the quality of the instrument to
38 the extent that each needs to be individually evaluated before reaching conclusions about its
39 effectiveness. Although results from meta-analytic studies can indicate the quality of different
40 selection methods in general, local validation studies are required to determine the effectiveness of
41 any given selection system.
42
43
44
45
46
47
48
49
50
51
52

53 **Implications for Theory**

54
55
56
57
58
59
60

1
2
3 A persistent problem with selection research relates to the question: what outcomes are we
4 trying to predict by using various selection methods? (56) To illustrate this ‘criterion problem’,
5
6
7 when exploring the association between conscientiousness and performance outcomes for
8
9
10 example, we find mixed results when examining outcomes relating to early exam performance in
11
12 medical school versus performance within clinical practice in later years. Furthermore, our review
13
14 also highlights that outcome measures used to evaluate selection methods most often focus on
15
16 indicators of attainment and maximal performance (e.g. medical school achievements,
17
18 performance in licensure exams) rather than indicators relating to clinical practice and typical (day
19
20 to day) performance in the job role.
21

22
23 Of the (few) longitudinal predictive validity studies available, often there lacks sufficient detail
24
25 regarding the target outcome variables with which to interpret results. In judging the evidence for
26
27 the relative accuracy of selection methods, there lacks a clear framework of outcome criteria with
28
29 which to interpret the research evidence and compare selection methods, both individually, and
30
31 within a selection system; future research should urgently address this gap in our understanding.
32
33

34 It is clear that indicators of competence during medical training and practice are likely to be
35
36 different at different points in a medical career - applicants are judged on multiple selection
37
38 criteria (depending on the specific role) which may include varying combinations of academic and
39
40 non-academic indicators of aptitude. A factor may be identified to be an important predictor for
41
42 undergraduate training, but may actually hinder some aspects of performance in clinical practice
43
44 (56, 60). Different selection methods may predict differently at different stages – for example, an
45
46 SJT may be less predictive of academic performance in the early years at medical school, but
47
48 significantly more predictive of performance outcomes once trainees enter clinical practice (27). A
49
50 major challenge within medicine is to integrate the research evidence to inform the design of
51
52 selection systems that are reliable and valid (and weighted appropriately) from undergraduate
53
54 selection through to selection for specialty training after many years of education, for both
55
56
57
58
59
60

1
2
3 academic and non-cognitive qualities. This requires a clearer, theoretically relevant taxonomy of
4
5 desirable outcomes variables which might range from academically-oriented variables such as
6
7 exam performance, through to variables relating to clinical practice and job performance
8
9 indicators as judged by supervisors, peers, and ideally, patients (e.g. multi-source feedback). As
10
11 such, there is a need for more theoretically driven, future-oriented, research aimed at identifying
12
13 what a “competent” doctor is at the various stages of training and practice, in order to move
14
15 towards crafting a unified taxonomy of performance indicators which may be used as markers in
16
17 short- and long-term predictive validity studies of selection methods.
18
19

20 21 **Implications for Practice**

22
23 A challenge in previous years has been to evaluate important non-cognitive attributes (e.g.
24
25 empathy, integrity) reliably at point of selection. Our review shows that SJTs and MMIs are more
26
27 valid predictors of inter- and intra-personal (non-academic) attributes than personal statements or
28
29 references. SJTs and MMIs are complementary: while SJTs can measure a broader range of
30
31 constructs efficiently as they can be machine marked, by contrast, MMIs are a face-to-face
32
33 encounter. Although expensive, structured interviews allow applicant responses to be probed
34
35 further and in more depth. Here, results from personality assessments could also add value when
36
37 used alongside a structured interview. An aim for future research and practice should be the design
38
39 and long-term evaluation of effective and scalable methods to assess non-academic attributes
40
41 accurately, and to explore the optimal combination of tools.
42
43
44

45
46 The picture at this point in time is less clear for aptitude tests and cognitive factors due to: the
47
48 large number of aptitude tests and the differences between currently-available aptitude tests; the
49
50 diverse outcome measures against which performance on aptitude tests is compared (to assess
51
52 validity - see the ‘criterion problem’ discussed above); the multiple ways in which aptitude tests
53
54 are implemented, and the mixed nature of the evidence on the effectiveness of aptitude testing.
55
56 There is also some evidence that some aptitude tests may favour certain types of candidates (43),
57
58
59
60

1
2
3 which may have unfavourable implications for fairness and widening access to medicine.
4
5 However, such a conclusion may be supported in future for specific aptitude tests and for specific
6
7 outcome measures, should further high quality research evidence become available.
8

9
10 Interpreting the breadth of currently available literature is challenging: while some
11
12 practitioners feel that there is insufficient evidence to evaluate selection methods, others argue that
13
14 there is so much evidence available that it is overwhelming to try to collate it to identify which
15
16 selection methods are the “best”. Challenges of interpreting and applying evidence of the relative
17
18 acceptability, cost-effectiveness, practical issues and effectiveness (including reliability and
19
20 validity) of selection methods include the lack of longitudinal data, no agreed-upon framework of
21
22 outcome criteria, and institutional differences (including available resources, their curriculum, and
23
24 differing philosophies of what a “high performing medical student” is considered to be by that
25
26 medical school). Indeed, Krieter and Axelson (1) acknowledge that the complexity of admissions
27
28 goals may also be an obstacle to evidence-based progress in medical school admissions, due to the
29
30 broad and frequently competing concerns regarding social justice, educational equality, healthcare
31
32 and political outcomes. Moreover, when judging the quality and effectiveness of selection
33
34 methods it is noteworthy that some criteria may compete with one another. For example, the
35
36 stakeholder acceptability for referees’ reports in selection is generally high, but the evidence for
37
38 validity is poor. Similarly, regarding other criteria, the evidence for validity of SCs is high but
39
40 they are relatively costly to implement. In this respect, when judging the quality and effectiveness
41
42 of different selection methods schools and employers may to choose to weight different features
43
44 depending on the context within which the selection system is operating. This review intends to
45
46 synthesise the literature for the reader to try to clarify the key understanding regarding the
47
48 strengths and limitations of each method, rather than prescribe a single, best approach. Ultimately,
49
50 the aim is to design efficient, acceptable and fair methods which are scalable for high volume use.
51
52
53
54
55
56
57
58
59
60

1
2
3 This review highlights that at present there is not sufficient evidence to suggest that any one of the
4
5 selection methods currently used meets all of these criteria.
6

7 The authors propose that a key implication for practice from the considerations above and for
8
9 the suggested research agenda outlined below, is the necessity for collaborative studies
10
11 internationally involving multiple sites to gather and analyse high quality, longitudinal data about
12
13 the effectiveness, cost efficiency, issues in implementation and stakeholder acceptability of
14
15 selection methods. In so doing there is an opportunity to gain practical, in-depth and long-term
16
17 knowledge about the relative efficiency of selection methods.
18
19

20 21 **Scoping a Future Research Agenda**

22 It is clear from our review that it is challenging to draw firm conclusions regarding the relative
23
24 strength of the different tools given the variety in quality and design of the currently available
25
26 research evidence: there is currently insufficient data on the effectiveness, procedural issues,
27
28 acceptability and cost effectiveness to propose a framework for international best practice in
29
30 medical selection methods. As such, here we outline a possible future research agenda which may
31
32 help to strengthen the evidence for each selection tool, in order to progress researchers' and
33
34 practitioners' knowledge towards a framework for best practice in medical selection methods.
35
36 Although the literature in selection methods is large, there exist many uncharted territories for
37
38 further research. There is a clear need for well-planned studies focusing on the *long-term follow-*
39
40 *up of students*, tracking students from admission through to assessments in more senior posts in
41
42 clinical practice, at point of licensure and beyond. This review clearly highlights the lack of
43
44 evidence available for schools and employers to use in making decisions about which selection
45
46 tools to use, in which combinations, and with what individual weightings apportioned to each tool
47
48 used.
49
50
51
52

53 Within the broader sphere of fairness issues in selection, more research is required
54
55 exploring *widening access and diversity* issues, whether it be race, ethnicity or social class, as this
56
57
58
59
60

1
2
3 remains a challenge within medical school admissions globally and it is becoming an increasingly
4
5 important consideration politically to reflect society within the healthcare professions. Indicators
6
7 of socio-demography pertinent to each country often reflect the same underpinning socio-
8
9 economic bias, which presents either a barrier to entry to study, or reduced chances of successful
10
11 application. The preceding literature review highlights a paucity of educational research of
12
13 sufficient quality and type to adequately assess the impact of a variety of selection tools upon
14
15 widening access robustly. For example, O'Neill and colleagues (162) found no significant effect
16
17 of selection method on social diversity in the medical student population, and suggest that it is
18
19 more important for widening access to attract a sufficiently diverse applicant pool than which
20
21 selection tool is used. Therefore, only tentative conclusions can be drawn. It is likely that some
22
23 selection tools are more sensitive to social bias than others but more definitive data is required.
24
25 For example, initial evaluation of SJTs at entry to medical school level confirms that applicants'
26
27 performance at testing does not follow the usual socio-economic trends as with tests of academic
28
29 attainment (163); further research is required to explain why this might be the case. There is also
30
31 initial evidence to suggest that MMIs may be equitable with regard to the demographic status of
32
33 applicants (134).
34
35
36
37

38 Reports which address aptitude tests, for example the UKCAT, have shown that institutions
39
40 whose selection policies favour using such a test tend to make more offers to applicants from
41
42 disadvantaged backgrounds, and that the aptitude test itself is less sensitive than traditional
43
44 measures of academic attainment to some socio-economic markers, such as school-type (47, 164).
45
46 Whilst traditional markers of prior educational attainment have been called the "academic
47
48 backbone" of medical education due to their highly predictive nature of subsequent performance
49
50 both at medical school and beyond, there is a need to explore how "contextual data" can be used to
51
52 allow the social and educational background of applicants to be taken into consideration alongside
53
54 their educational achievement. Prior academic attainment is clearly still an important component
55
56
57
58
59
60

1
2
3 of the medical selection process but care must be taken to ensure that this is done in such a way to
4
5 ensure that it is not a barrier to candidates from disadvantaged groups.
6

7 A key criticism of selection research is that there is a distinct lack of *theory-driven studies* in
8
9 terms of issues related to validity and what constructs are being measured and, more broadly,
10
11 acknowledging contemporary models of adult intellectual development and skill acquisition, or
12
13 which attempt to integrate cognitive and non-cognitive factors (160, 161). The term “non-
14
15 cognitive” is in itself problematic: future research must also look towards more theoretical
16
17 underpinnings, drawing on not just psychometric approaches but also theoretical models of adult
18
19 intellectual functioning, personality, values, and individual differences. For example, there has
20
21 been little previous research exploring how to assess values as part of recruitment to the healthcare
22
23 professions, yet compassion and benevolence are important for any healthcare professional to
24
25 ensure the provision of high quality care and patient outcomes, and so new research literature in
26
27 this area is now emerging (165). Only by exploring the theoretical underpinnings will research in
28
29 selection progress to enable a richer understanding of how personality, aptitude, interest, values
30
31 and motivation interact to define areas of competence and career choice.
32
33
34
35

36 In summary, we propose the following priorities for a future research agenda, in order to enable
37
38 schools and employers to make evidence-based decisions about which selection tools and why:
39

- 40 • Longitudinal research exploring predictive validity and following students throughout
41
42 the course of their career within education, training and practice
- 43 • Research enabling greater understanding of how selection tools may impact on
44
45 widening access and diversity agendas
- 46 • Theoretically driven studies of the construct validity of both cognitive and non-
47
48 cognitive oriented selection methods and also selection systems, in order to understand
49
50 what we are assessing for in both the short- and long-term
51
52
53
54
55

56 **Strengths and Limitations**

57
58
59
60

1
2
3 A key strength of our review is that we collate and synthesise the breadth of research evidence
4 over the last 15 years in order to draw conclusions regarding five key evaluation criteria (*type of*
5 *evidence; effectiveness; procedural issues; cost-effectiveness; and stakeholder acceptance*)
6 regarding medical selection methods. We also identify current gaps in understanding and theory,
7 and outline a future research agenda which aims to address these areas.
8
9
10
11
12

13
14 Attempting to summarise our conclusions of the large number of studies reviewed in Table 4
15 naturally runs the risk of simplifying some of the intricacies of the studies and the nuances of their
16 findings. The authors therefore encourage the reader to consider the original source should they
17 wish to gain a fuller picture of each study's context, rationale, methodology and findings.
18 Nonetheless, the authors feel that Table 4 provides a valuable resource for the reader to identify
19 key papers and navigate the sizeable and diverse literature base.
20
21
22
23
24
25
26
27
28

29 **Acknowledgements**

30
31 *Our thanks to the General Medical Council (GMC) of the UK for commissioning us to carry out*
32 *an initial rapid review of the literature on selection and widening access to medicine in 2013. Our*
33 *thanks also to Medical Schools Council (MSC) of the UK for commissioning us to conduct a*
34 *significantly updated review of selection methods in 2014; funding for which was provided from*
35 *Health Education England (HEE) and the Office for Fair Access (OFFA). Working together both*
36 *of these projects encouraged us to further develop our ideas and produce an updated systematic*
37 *review for publication in 2015. Our thanks to those who contributed to the original project funded*
38 *by the GMC, notably Professor John McLachlan and Ms Emma Dunlop.*
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

1. Kreiter CD, Axelson RD. A perspective on medical school admission research and practice over the last 25 years. *Teaching and Learning in Medicine*. 2013;25:32-8015 (Electronic).
2. Coates H. Establishing the criterion validity of the Graduate Medical School Admissions Test (GAMSAT). *Medical Education*. 2008;42(10):999-1006.
3. Trost G, Nauels HU, Klieme E. The relationship between different criteria for admission to medical school and student success. *Assessment in Education: Principles, Policy & Practice*. 1998;5(2):247-54.
4. Halpenny D, Cadoo K, Halpenny M, Burke J, Torreggiani WC. The Health Professions Admission Test (HPAT) score and leaving certificate results can independently predict academic performance in medical school: do we need both tests? *Irish Medical Journal*. 2010;103(10):300-2.
5. Patterson F, Ferguson E. Selection for medical education and training: Wiley Online Library; 2010.
6. Ferguson E, James D, Madeley L. Factors associated with success in medical school: systematic review of the literature. *British Medical Journal*. 2002;324(7343):952-7.
7. Milburn A. University challenge: how higher education can advance social mobility. 2012.
8. Prideaux D, Roberts C, Eva K, Centeno A, McCrorie P, McManus C, et al. Assessment for selection for the health care professions and specialty training: consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher*. 2011;33(3):215-23.
9. Petticrew M, Roberts H. Systematic reviews in the social sciences: A practical guide: John Wiley & Sons; 2008.
10. Callahan CA, Hojat M, Veloski J, Erdmann JB, Gonnella JS. The predictive validity of three versions of the MCAT in relation to performance in medical school, residency, and licensing examinations: A longitudinal study of 36 classes of Jefferson Medical College. *Academic Medicine*. 2010;85(6):980-7.
11. Dunleavy DM, Kroopnick MH, Dowd KW, Searcy CA, Zhao X. The predictive validity of the MCAT exam in relation to academic performance through medical school: a national cohort study of 2001-2004 matriculants. *Academic Medicine: Journal of the Association of American Medical Colleges*. 2013;88(5):666-71.
12. Elam CL, Stratton TD, Scott KL, Wilson JF, Lieber A. Review, deliberation, and voting: a study of selection decisions in a medical school admission committee. *Teaching and Learning in Medicine*. 2002;14(2):98-103.
13. Peskun C, Detsky A, Shandling M. Effectiveness of medical school admissions criteria in predicting residency ranking four years later. *Medical Education*. 2007;41(1):57-64.
14. Puddey IB, Mercer A. Predicting academic outcomes in an Australian graduate entry medical programme. *BMC Medical Education*. 2014;14:31.
15. Edwards D, Friedman T, Pearce J. Same admissions tools, different outcomes: a critical perspective on predictive validity in three undergraduate medical schools. *BMC Medical Education*. 2013;13:173.
16. Poole P, Shulruf B, Rudland J, Wilkinson T. Comparison of UMAT scores and GPA in prediction of performance in medical school: a national study. *Medical Education*. 2012;46(2):163-71.
17. Husbands A, Mathieson A, Dowell J, Cleland J, MacKenzie R. Predictive validity of the UK clinical aptitude test in the final years of medical school: a prospective cohort study. *BMC Medical Education*. 2014;14:88.
18. McManus IC, Dewberry C, Nicholson S, Dowell JS. The UKCAT-12 study: educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a cross-sectional collaborative study of 12 UK medical schools. *BMC Medicine*. 2013;11:244.
19. Sartania N, McClure JD, Sweeting H, Browitt A. Predictive power of UKCAT and other pre-admission measures for performance in a medical school in Glasgow: a cohort study. *BMC Medical Education*. 2014;14:116.
20. Wright SR, Bradley PM. Has the UK Clinical Aptitude Test improved medical student selection? *Medical Education*. 2010;44(11):1069-76.

21. Bell JF. The case against the BMAT: Not Withering but Withered? *British Medical Journal*. 2005;331:555.
22. Emery JL, Bell JF, Vidal Rodeiro CL. The BioMedical Admissions Test for medical student selection: Issues of fairness and bias. *Medical Teacher*. 2011;33(1):62-71.
23. Albishri JA, Aly SM, Alnemary Y. Admission criteria to Saudi medical schools. Which is the best predictor for successful achievement? *Saudi Medical Journal*. 2012;33(11):1222-6.
24. Moore EJ, Price DL, Van Abel KM, Carlson ML. Still under the microscope: Can a surgical aptitude test predict otolaryngology resident performance? *The Laryngoscope*. 2014.
25. Donnon T, Paolucci EO, Violato C. The predictive validity of the MCAT for medical school performance and medical board licensing examinations: A meta-analysis of the published research. *Academic Medicine*. 2007;82(1):100-6.
26. Yates J, James D. The value of the UK Clinical Aptitude Test in predicting pre-clinical performance: a prospective cohort study at Nottingham Medical School. *BMC Medical Education*. 2010;10(1):1-9.
27. Wilkinson D, Zhang J, Byrne GJ, Luke H, Ozolins IZ, Parker MH, Peterson RF. Medical school selection criteria and the prediction of academic performance. *Medical Journal of Australia*. 2008;188(6):349-54.
28. Griffin B, Yeomans ND, Wilson IG. Students coached for an admission test perform less well throughout a medical course. *Internal Medicine Journal*. 2013;43(8):927-32.
29. Laurence CO, Zajac IT, Lorimer M, Turnbull DA, Sumner KE. The impact of preparatory activities on medical school selection outcomes: a cross-sectional survey of applicants to the University of Adelaide Medical School in 2007. *BMC Medical Education*. 2013;13:159.
30. Poole P, Shulruf B. Shaping the future medical workforce: take care with selection tools. *Journal of Primary Health Care*. 2013;5(4):269-75.
31. Puddey IB, Mercer A, Andrich D, Styles I. Practice effects in medical school entrance testing with the undergraduate medicine and health sciences admission test (UMAT). *BMC Medical Education*. 2014;14:48.
32. Simpson PL, Scicluna HA, Jones PD, Cole AM, O'Sullivan AJ, Harris PG, Velan G, McNeil HP. Predictive validity of a new integrated selection process for medical school admission. *BMC Medical Education*. 2014;14:86.
33. Wilkinson D, Zhang J, Parker M. Predictive validity of the Undergraduate Medicine and Health Sciences Admission Test for medical students' academic performance. *Medical Journal of Australia*. 2011;194(7):341-4.
34. McManus IC, Ferguson E, Wakeford R, Powis D, James D. Predictive validity of the Biomedical Admissions Test: An evaluation and case study. *Medical Teacher*. 2011;33(1):53-7.
35. McManus IC, Ferguson E, Wakeford R, Powis D, James D. Response to comments by Emery and Bell, *Medical Teacher* 33(1): (this issue). *Medical Teacher*. 2011;33(1):60-1.
36. Al-Rukban MO, Munshi FM, Abdulghani HM, Al-Hoqaill I. The ability of the pre-admission criteria to predict performance in a Saudi medical school. *Saudi Medical Journal*. 2010;31(5):560-4.
37. Kraft HG, Lamina C, Kluckner T, Wild C, Prodingner WM. Paradise lost or paradise regained? Changes in admission system affect academic performance and drop-out rates of medical students. *Medical Teacher*. 2013;35(5):e1123-9.
38. Adam J, Dowell J, Greatrix R. Use of UKCAT scores in student selection by U.K. medical schools, 2006-2010. *BMC Medical Education*. 2011;11:98.
39. Albanese MA, Farrell P, Dottl S. Statistical criteria for setting thresholds in medical school admissions. *Advances in Health Sciences Education: Theory and Practice*. 2005;10(2):89-103.
40. Hissbach J, Klusmann D, Hampe W. Dimensionality and predictive validity of the HAM-Nat, a test of natural sciences for medical school admission. *BMC Medical Education*. 2011;11(1):1-11.
41. Zhao X, Oppler S, Dunleavy D, Kroopnick M. Validity of four approaches of using repeaters' MCAT scores in medical school admissions to predict USMLE Step 1 total scores. *Academic Medicine: Journal of the Association of American Medical Colleges*. 2010;85(10 Suppl):S64-7.
42. Cleland JA, French FH, Johnston PW. A mixed-methods study identifying and exploring medical students' views of the UKCAT. *Medical Teacher*. 2011;33(3):244-9.
43. Aldous CJ, Leeder SR, Price J, Sefton AE, Teubner JK. A selection test for Australian graduate-entry medical schools. *Medical Journal of Australia*. 1997;166(5):247-50.

- 1
- 2
- 3 44. Griffin B, Harding DW, Wilson IG, Yeomans ND. Does practice make perfect? The effect of
- 4 coaching and retesting on selection tests used for admission to an Australian medical school. *Medical*
- 5 *Journal of Australia*. 2008;189(5):270-3.
- 6 45. Lambe P, Waters C, Bristow D. The UK Clinical Aptitude Test: is it a fair test for selecting
- 7 medical students? *Medical Teacher*. 2012;34(8):e557-65.
- 8 46. Winegarden B, Glaser D, Schwartz A, Kelly C. MCAT Verbal Reasoning score: less
- 9 predictive of medical school performance for English language learners. *Medical Education*.
- 10 2012;46(9):878-86.
- 11 47. Tiffin PA, McLachlan JC, Webster L, Nicholson S. Comparison of the sensitivity of the
- 12 UKCAT and A Levels to sociodemographic characteristics: a national study. *BMC Medical*
- 13 *Education*. 2014;14:7.
- 14 48. Griffin BN, Wilson IG. Interviewer bias in medical student selection. *Medical Journal of*
- 15 *Australia*. 2010;193(6):343-6.
- 16 49. O'Flynn S, Fitzgerald T, Mills A. Modelling the impact of old and new mechanisms of entry
- 17 and selection to medical school in Ireland: who gets in? *Irish Journal of Medical Science*.
- 18 2013;182(3):421-7.
- 19 50. Bhatti MA, Anwar M. Does entry test make any difference on the future performance of
- 20 medical students? *Journal of Pakistan Medical Association*. 2012;62(7):664-8.
- 21 51. Cohen-Schotanus J, Muijtjens AM, Reinders JJ, Agsteribbe J, van Rossum HJ, van der
- 22 Vleuten CP. The predictive validity of grade point average scores in a partial lottery medical school
- 23 admission system. *Medical Education*. 2006;40(10):1012-9.
- 24 52. Kreiter CD, Kreiter Y. A validity generalization perspective on the ability of undergraduate
- 25 GPA and the medical college admission test to predict important outcomes. *Teaching and Learning in*
- 26 *Medicine*. 2007;19(2):95-100.
- 27 53. Lumb AB, Vail A. Comparison of academic, application form and social factors in predicting
- 28 early performance on the medical course. *Medical Education*. 2004;38(9):1002-5.
- 29 54. Luqman M. Relationship of academic success of medical students with motivation and pre-
- 30 admission grades. *Journal of the College of Physicians and Surgeons--Pakistan: JCPSP*.
- 31 2013;23(1):31-6.
- 32 55. McManus IC, Smithers E, Partridge P, Keeling A, Fleming PR. A levels and intelligence as
- 33 predictors of medical careers in UK doctors: 20 year prospective study. *British Medical Journal*.
- 34 2003;327(7407):139-42.
- 35 56. Ferguson E, Semper H, Yates J, Fitzgerald JE, Skatova A, James D. The 'dark side' and
- 36 'bright side' of personality: when too much conscientiousness and too little anxiety are detrimental
- 37 with respect to the acquisition of medical knowledge and skill. *PLoS One*. 2014;9(2):e88606.
- 38 57. McManus IC, Dewberry C, Nicholson S, Dowell JS, Woolf K, Potts HWW. Construct-level
- 39 predictive validity of educational attainment and intellectual aptitude tests in medical student
- 40 selection: meta-regression of six UK longitudinal studies. *BMC Medicine*. 2013;11:243.
- 41 58. O'Neill L, Hartvigsen J, Wallstedt B, Korsholm L, Eika B. Medical school dropout--testing at
- 42 admission versus selection by highest grades as predictors. *Medical Education*. 2011;45(11):1111-20.
- 43 59. Urlings-Strop LC, Stegers-Jager KM, Stijnen T, Themmen AP. Academic and non-academic
- 44 selection criteria in predicting medical school performance. *Medical Teacher*. 2013;35(6):497-502.
- 45 60. Ferguson E, James D, O'Hehir F, Sanders A, McManus IC. Pilot study of the roles of
- 46 personality, references, and personal statements in relation to performance over the five years of a
- 47 medical degree. *British Medical Journal*. 2003;326(7386):429-32.
- 48 61. Tektas OY, Fiessler C, Mayr A, Neuhuber W, Paulsen F. Correlation of high school exam
- 49 grades with study success at a German medical school. *Journal of Contemporary Medical Education*.
- 50 2013;1(3):157-62.
- 51 62. McManus IC, Powis DA, Wakeford R, Ferguson E, James D, Richards P. Intellectual aptitude
- 52 tests and A levels for selecting UK school leaver entrants for medical school. *British Medical Journal*.
- 53 2005;331(7516):555-9.
- 54 63. McManus IC, Woolf K, Dacre J. Even one star at A level could be "too little, too late" for
- 55 medical student selection. *BMC Medical Education*. 2008;8(1):16.
- 56 64. Oosterveld P, ten Cate O. Generalizability of a study sample assessment procedure for
- 57 entrance selection for medical school. *Medical Teacher*. 2004;26(7):635-9.
- 58
- 59
- 60

- 1
2
3 65. Ferguson E, Sanders A, O'Hehir F, James D. Predictive validity of personal statements and
4 the role of the five-factor model of personality in relation to medical training. *Journal of Occupational*
5 *and Organizational Psychology*. 2000;73(3):321-44.
- 6 66. Husbands A, Dowell J. Predictive validity of the Dundee multiple mini-interview. *Medical*
7 *Education*. 2013;47(7):717-25.
- 8 67. White J, Brownell K, Lemay JF, Lockyer JM. "What do they want me to say?" The hidden
9 curriculum at work in the medical school selection process: a qualitative study. *BMC medical*
10 *education*. 2012;12:17.
- 11 68. White JS, Lemay JF, Brownell K, Lockyer J. "A chance to show yourself" - how do
12 applicants approach medical school admission essays? *Medical Teacher*. 2011;33(10):e541-8.
- 13 69. Elam CL, Johnson MM. The effect of a rolling admission policy on a medical school's
14 selection of applicants. *Academic Medicine: Journal of the Association of American Medical*
15 *Colleges*. 1997;72(7):644-6.
- 16 70. Hanson MD, Dore KL, Reiter HI, Eva KW. Medical school admissions: revisiting the
17 veracity and independence of completion of an autobiographical screening tool. *Academic Medicine :*
18 *Journal of the Association of American Medical Colleges*. 2007;82(10 Suppl):S8-S11.
- 19 71. Parry J, Mathers J, Stevens A, Parsons A, Lilford R, Spurgeon P, Thomas H.. Admissions
20 processes for five year medical courses at English schools: review. *British Medical Journal*.
21 2006;332(7548):1005-9.
- 22 72. Kumwenda B, Dowell J, Husbands A. Is embellishing UCAS personal statements accepted
23 practice in applications to medicine and dentistry? *Medical Teacher*. 2013;35(7):599-603.
- 24 73. Benbassat J, Baumal R. Uncertainties in the selection of applicants for medical school.
25 *Advances in Health Sciences Education: Theory and Practice*. 2007;12(4):509-21.
- 26 74. Poole PJ, Moriarty HJ, Wearn AM, Wilkinson TJ, Weller JM. Medical student selection in
27 New Zealand: looking to the future. *The New Zealand Medical Journal*. 2009;122(1306):88-100.
- 28 75. Stedman JM, Hatch JP, Schoenfeld LS. Letters of recommendation for the predoctoral
29 internship in medical schools and other settings: do they enhance decision making in the selection
30 process? *Journal of Clinical Psychology in Medical Settings*. 2009;16(4):339-45.
- 31 76. Rostom H, Watson R, Leaver L. Situational judgement tests: the role of coaching. *Medical*
32 *Education*. 2013;47(2):219-.
- 33 77. Cabrera MAM, Nguyen NT. Situational Judgment Tests: A Review of Practice and
34 Constructs Assessed. *International Journal of Selection and Assessment*. 2001;9(1-2):103-13.
- 35 78. Christian MS, Edwards BD, Bradley JC. Situational Judgement Tests: Constructs assessed
36 and a meta-analysis of their criterion-related validities. *Personnel Psychology*. 2010;63(1):83-117.
- 37 79. Hansel M, Klupp S, Graupner A, Dieter P, Koch T. Dresden Faculty selection procedure for
38 medical students: what impact does it have, what is the outcome? *GMS Zeitschrift für Medizinische*
39 *Ausbildung*. 2010;27(2):Doc25.
- 40 80. Lievens F. Adjusting medical school admission: assessing interpersonal skills using
41 situational judgement tests. *Medical Education*. 2013;47(2):182-9.
- 42 81. Lievens F, Buyse T, Sackett PR. The operational validity of a video-based situational
43 judgment test for medical college admissions: illustrating the importance of matching predictor and
44 criterion construct domains. *The Journal of Applied Psychology*. 2005;90(3):442-52.
- 45 82. Lievens F, Peeters H, Schollaert E. Situational judgment tests: A review of recent research.
46 *Personnel Review*. 2008;37(4):426-41.
- 47 83. Patterson F, Carr V, Zibarras L, Burr B, Berkin L, Plint S, et al. New machine-marked tests
48 for selection into core medical training: evidence from two validation studies. *Clinical Medicine*
49 *(London, England)*. 2009;9(5):417-20.
- 50 84. Libbrecht N, Lievens F, Carette B, Cote S. Emotional intelligence predicts success in medical
51 school. *Emotion (Washington, DC)*. 2014;14(1):64-73.
- 52 85. Patterson F, Ashworth V, Kerrin M, O'Neill P. Situational judgement tests represent a
53 measurement method and can be designed to minimise coaching effects. *Medical Education*.
54 2013;47(2):220-1.
- 55 86. Lievens F, Sackett PR. Situational judgment tests in high-stakes settings: issues and strategies
56 with generating alternate forms. *The Journal of Applied Psychology*. 2007;92(4):1043-55.
- 57
58
59
60

- 1
- 2
- 3 87. McDaniel MA, Hartman NS, Whetzel DL, Grubb WL. Situational judgement tests, response
- 4 instructions, and validity: A meta-analysis. *Personnel Psychology*. 2007;60(1):63-91.
- 5 88. Cullen MJ, Sackett PR, Lievens F. Threats to the Operational Use of Situational Judgment
- 6 Tests in the College Admission Process. *International Journal of Selection and Assessment*.
- 7 2006;14(2):142-55.
- 8 89. Chan D, Schmitt N. Situational Judgment and Job Performance. *Human Performance*.
- 9 2002;15(3):233-54.
- 10 90. Koczwara A, Patterson F, Zibarras L, Kerrin M, Irish B, Wilkinson M. Evaluating cognitive
- 11 ability, knowledge tests and situational judgement tests for postgraduate selection. *Medical Education*.
- 12 2012;46(4):399-408.
- 13 91. Plint S, Patterson F. Identifying critical success factors for designing selection processes into
- 14 postgraduate specialty training: the case of UK general practice. *Postgraduate Medical Journal*.
- 15 2010;86(1016):323-7.
- 16 92. Ahmed H, Rhydderch M, Matthews P. Can knowledge tests and situational judgement tests
- 17 predict selection centre performance? *Medical Education*. 2012;46(8):777-84.
- 18 93. Clevenger J, Pereira GM, Wiechmann D, Schmitt N, Harvey VS. Incremental validation of
- 19 situational judgment tests. *The Journal of Applied Psychology*. 2001;86(3):410-7.
- 20 94. O'Connell MS, Hartman NS, McDaniel MA, Grubb WL, Lawrence A. Incremental Validity
- 21 of Situational Judgment Tests for Task and Contextual Job Performance. *International Journal of*
- 22 *Selection and Assessment*. 2007;15(1):19-29.
- 23 95. Patterson F, Baron H, Carr V, Plint S, Lane P. Evaluation of three short-listing methodologies
- 24 for selection into postgraduate training in general practice. *Medical Education*. 2009;43(1):50-7.
- 25 96. Haight SJ, Chibnall JT, Schindler DL, Slavin SJ. Associations of medical student personality
- 26 and health/wellness characteristics with their medical school performance across the curriculum.
- 27 *Academic Medicine: Journal of the Association of American Medical Colleges*. 2012;87(4):476-85.
- 28 97. Hojat M, Erdmann JB, Gonnella JS. Personality assessments and outcomes in medical
- 29 education and the practice of medicine: AMEE Guide No. 79. *Medical Teacher*. 2013;35(7):e1267-
- 30 e301
- 31 98. Lievens F, Coetsier P, De Fruyt F, De Maeseneer J. Medical students' personality
- 32 characteristics and academic performance: a five-factor model perspective. *Medical Education*.
- 33 2002;36(11):1050-6.
- 34 99. Knights J, Kennedy B. Medical school selection: Screening for dysfunctional tendencies.
- 35 *Medical Education*. 2006;40:1058-64.
- 36 100. Knights JA, Kennedy BJ. Medical school selection: impact of dysfunctional tendencies on
- 37 academic performance. *Medical Education*. 2007;41(4):362-8.
- 38 101. Chan-Ob T, Boonyanaruthee V. Medical student selection: which matriculation scores and
- 39 personality factors are important? *Journal of the Medical Association of Thailand - Chotmaihet*
- 40 *Thangphaet*. 1999;82(6):604-10.
- 41 102. Carrothers RM, Gregory SW, Jr., Gallagher TJ. Measuring emotional intelligence of medical
- 42 school applicants. *Academic Medicine: Journal of the Association of American Medical Colleges*.
- 43 2000;75(5):456-63.
- 44 103. Edwards JC, Elam CL, Wagoner NE. An admission model for medical schools. *Academic*
- 45 *medicine : journal of the Association of American Medical Colleges*. 2001;76(12):1207-12.
- 46 104. Bore M, Munro D, Powis D. A comprehensive model for the selection of medical students.
- 47 *Medical Teacher*. 2009;31(12):1066-72.
- 48 105. Carr SE. Emotional intelligence in medical students: does it correlate with selection
- 49 measures? *Medical Education*. 2009;43(11):1069-77.
- 50 106. Lin DT, Kannappan A, Lau JN. The assessment of emotional intelligence among candidates
- 51 interviewing for general surgery residency. *Journal of Surgical Education*. 2013;70(4):514-21.
- 52 107. Leddy JJ, Moineau G, Puddester D, Wood TJ, Humphrey-Murto S. Does an emotional
- 53 intelligence test correlate with traditional measures used to determine medical school admission?
- 54 *Academic Medicine: Journal of the Association of American Medical Colleges*. 2011;86(10
- 55 *Suppl*):S39-41.
- 56
- 57
- 58
- 59
- 60

108. Brannick MT, Grichanik M, Nazian S, Wahi M, Goldin S. Emotional Intelligence and Medical School Performance: A Prospective Multivariate Study. *Medical Sciences Education*. 2013;23(4):628-36.
109. Cherry MG, Fletcher I, O'Sullivan H, Dornan T. Emotional intelligence in medical education: a critical review. *Medical Education*. 2014;48(5):468-78.
110. Lievens F, Ones DS, Dilchert S. Personality scale validities increase throughout medical school. *The Journal of Applied Psychology*. 2009;94(6):1514-35.
111. Jerant A, Griffin E, Rainwater J, Henderson M, Sousa F, Bertakis KD, et al. Does applicant personality influence multiple mini-interview performance and medical school acceptance offers? *Academic Medicine: Journal of the Association of American Medical Colleges*. 2012;87(9):1250-9.
112. Powis DA, Rolfe I. Selection and performance of medical students at Newcastle. *New South Wales Education for Health*. 1998;11:15-23.
113. Ann Courneya C, Wright K, Frinton V, Mak E, Schulzer M, Pachev G. Medical student selection: choice of a semi-structured panel interview or an unstructured one-on-one interview. *Medical Teacher*. 2005;27(6):499-503.
114. Donnon T, Oddone-Paolucci E, Violato C. A predictive validity study of medical judgment vignettes to assess students' noncognitive attributes: a 3-year prospective longitudinal study. *Medical Teacher*. 2009;31(4):e148-55.
115. Donnon T, Paolucci EO. A generalizability study of the medical judgment vignettes interview to assess students' noncognitive attributes for medical school. *BMC Medical Education*. 2008;8:58.
116. Elam CL, Studts JL, Johnson MMS. Prediction of medical school performance: Use of admission interview report narratives. *Teaching and Learning in Medicine*. 1997;9(3):181-5.
117. Kleshinski J, Shriner C, Khuder SA. The use of professionalism scenarios in the medical school interview process: faculty and interviewee perceptions. *Medical Education Online*. 2008;13:2.
118. Patrick LE, Altmaier EM, Kuperman S, Ugolini K. A Structured Interview for Medical School Admission, Phase 1: Initial Procedures and Results. *Academic Medicine*. 2001;76(1):66-71.
119. Rahbar MH, Vellani C, Sajjan F, Zaidi AA, Akbarali L. Predictability of medical students' performance at the Aga Khan University from admission test scores, interview ratings and systems of education. *Medical Education*. 2001;35(4):374-80.
120. VanSusteren TJ, Suter E, Romrell LJ, Lanier L, Hatch RL. Do Interviews Really Play an Important Role in the Medical School Selection Decision? *Teaching and Learning in Medicine*. 1999;11(2):66-74.
121. Basco WT, Jr., Gilbert GE, Chessman AW, Blue AV. The ability of a medical school admission process to predict clinical performance and patients' satisfaction. *Academic Medicine: Journal of the Association of American Medical Colleges*. 2000;75(7):743-7.
122. Basco WT, Lancaster C, Carey ME, Gilbert GE, Blue AV, editors. The medical school applicant interview predicts performance on a fourth-year clinical practice examination. *Pediatric Research*; 2004.
123. Basco WT, Jr., Lancaster CJ, Gilbert GE, Carey ME, Blue AV. Medical school application interview score has limited predictive validity for performance on a fourth year clinical practice examination. *Advances in Health Sciences Education: Theory and Practice*. 2008;13(2):151-62.
124. Fan AP, Tsai TC, Su TP, Kosik RO, Morisky DE, Chen CH, et al. A longitudinal study of the impact of interviews on medical school admissions in Taiwan. *Evaluation & the Health Professions*. 2010;33(2):140-63.
125. Kreiter C, Yin P, Solow C, Brennan R. Investigating the Reliability of the Medical School Admissions Interview. *Advances in Health Sciences Education*. 2004;9(2):147-59.
126. Streyffeler L, Altmaier EM, Kuperman S, Patrick LE. Development of a Medical School Admissions Interview Phase 2: Predictive Validity of Cognitive and Non-Cognitive Attributes. *Medical Education Online*. 2009;10.
127. Casey M, Wilkinson D, Fitzgerald J, Eley D, Connor J. Clinical communication skills learning outcomes among first year medical students are consistent irrespective of participation in an interview for admission to medical school. *Medical Teacher*. 2014;36(7):640-2.
128. Eva KW, Macala C. Multiple mini-interview test characteristics: 'tis better to ask candidates to recall than to imagine. *Medical Education*. 2014;48(6):604-13.

- 1
2
3 129. Dore KL, Kreuger S, Ladhani M, Rolfson D, Kurtz D, Kulasegaram K, et al. The reliability
4 and acceptability of the Multiple Mini-Interview as a selection instrument for postgraduate
5 admissions. *Academic Medicine: Journal of the Association of American Medical Colleges*.
6 2010;85(10 Suppl):S60-3.
- 7 130. Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-
8 interview. *Medical Education*. 2004;38(3):314-26.
- 9 131. Hofmeister M, Lockyer J, Crutcher R. The acceptability of the multiple mini interview for
10 resident selection. *Family Medicine*. 2008;40(10):734-40.
- 11 132. O'Brien A, Harvey J, Shannon M, Lewis K, Valencia O. A comparison of multiple mini-
12 interviews and structured interviews in a UK setting. *Medical Teacher*. 2011;33(5):397-402.
- 13 133. Roberts C, Walton M, Rothnie I, Crossley J, Lyon P, Kumar K, et al. Factors affecting the
14 utility of the multiple mini-interview in selecting candidates for graduate-entry medical school.
15 *Medical education*. 2008;42(4):396-404.
- 16 134. Campagna-Vaillancourt M, Manoukian J, Razack S, Nguyen LH. Acceptability and reliability
17 of multiple mini interviews for admission to otolaryngology residency. *The Laryngoscope*.
18 2014;124(1):91-6.
- 19 135. Hissbach JC, Sehner S, Harendza S, Hampe W. Cutting costs of multiple mini-interviews -
20 changes in reliability and efficiency of the Hamburg medical school admission test between two
21 applications. *BMC Medical Education*. 2014;14:54.
- 22 136. Eva KW, Reiter HI, Rosenfeld J, Norman GR. The ability of the multiple mini-interview to
23 predict preclerkship performance in medical school. *Academic Medicine: Journal of the Association*
24 *of American Medical Colleges*. 2004;79(10 Suppl):S40-2.
- 25 137. Eva KW, Reiter HI, Trinh K, Wasi P, Rosenfeld J, Norman GR. Predictive validity of the
26 multiple mini-interview for selecting medical trainees. *Medical Education*. 2009;43(8):767-75.
- 27 138. Hofmeister M, Lockyer J, Crutcher R. The multiple mini-interview for selection of
28 international medical graduates into family medicine residency education. *Medical Education*.
29 2009;43(6):573-9.
- 30 139. Reiter HI, Eva KW, Rosenfeld J, Norman GR. Multiple mini-interviews predict clerkship and
31 licensing examination performance. *Medical Education*. 2007;41(4):378-84.
- 32 140. Rosenfeld JM, Reiter HI, Trinh K, Eva KW. A cost efficiency comparison between the
33 multiple mini-interview and traditional admissions interviews. *Advances in Health Sciences*
34 *Education: Theory and Practice*. 2008;13(1):43-58.
- 35 141. Hopson LR, Burkhardt JC, Stansfield RB, Vohra T, Turner-Lawrence D, Losman ED. The
36 multiple mini-interview for emergency medicine resident selection. *The Journal of Emergency*
37 *Medicine*. 2014;46(4):537-43.
- 38 142. Pau A, Jeevaratnam K, Chen YS, Fall AA, Khoo C, Nadarajah VD. The Multiple Mini-
39 Interview (MMI) for student selection in health professions training - a systematic review. *Medical*
40 *Teacher*. 2013;35(12):1027-41.
- 41 143. Axelson R, Kreiter C, Ferguson K, Solow C, Huebner K. Medical School Preadmission
42 Interviews: Are Structured Interviews More Reliable Than Unstructured Interviews? *Teaching and*
43 *Learning in Medicine*. 2010;22(4):241-5.
- 44 144. Kumar K, Roberts C, Rothnie I, du Fresne C, Walton M. Experiences of the multiple mini-
45 interview: a qualitative analysis. *Medical Education*. 2009;43(4):360-7.
- 46 145. Quintero AJ, Segal LS, King TS, Black KP. The Personal Interview: Assessing the Potential
47 for Personality Similarity to Bias the Selection of Orthopaedic Residents. *Academic Medicine*.
48 2009;84(10):1364-72 10.097/ACM.0b013e3181b6a9af.
- 49 146. Razack S, Faremo S, Drolet F, Snell L, Wiseman J, Pickering J. Multiple mini-interviews
50 versus traditional interviews: stakeholder acceptability comparison. *Medical Education*.
51 2009;43(10):993-1000.
- 52 147. McManus IC, Richards P, Winder BC. Do UK Medical School Applicants Prefer
53 Interviewing to Non-Interviewing Schools? *Advances in Health Sciences Education*. 1999;4(2):155-
54 65.
- 55 148. Dowell J, Lynch B, Till H, Kumwenda B, Husbands A. The multiple mini-interview in the
56 U.K. context: 3 years of experience at Dundee. *Medical Teacher*. 2012;34(4):297-304.
- 57
58
59
60

149. Humphrey S, Dowson S, Wall D, Diwakar V, Goodyear HM. Multiple mini-interviews: opinions of candidates and interviewers. *Medical Education*. 2008;42(2):207-13.
150. Sebok SS, Luu K, Klinger DA. Psychometric properties of the multiple mini-interview used for medical admissions: findings from generalizability and Rasch analyses. *Advances in Health Sciences Education: Theory and Practice*. 2014;19(1):71-84.
151. Dodson M, Crotty B, Prideaux D, Carne R, Ward A, de Leeuw E. The multiple mini-interview: how long is long enough? *Medical Education*. 2009;43(2):168-74.
152. Tiller D, O'Mara D, Rothnie I, Dunn S, Lee L, Roberts C. Internet-based multiple mini-interviews for candidate selection for graduate entry programmes. *Medical Education*. 2013;47(8):801-10.
153. Gafni N, Moshinsky A, Eisenberg O, Zeigler D, Ziv A. Reliability estimates: behavioural stations and questionnaires in medical school admissions. *Medical education*. 2012;46(3):277-88.
154. ten Cate O, Smal K. Educational assessment center techniques for entrance selection in medical school. *Academic Medicine: Journal of the Association of American Medical Colleges*. 2002;77(7):737.
155. Ziv A, Rubin O, Moshinsky A, Gafni N, Kotler M, Dagan Y, et al. MOR: a simulation-based assessment centre for evaluating the personal and interpersonal qualities of medical school candidates. *Medical Education*. 2008;42(10):991-8.
156. Gale TC, Roberts MJ, Sice PJ, Langton JA, Patterson FC, Carr AS, et al. Predictive validity of a selection centre testing non-technical skills for recruitment to training in anaesthesia. *British Journal of Anaesthesia*. 2010;105(5):603-9.
157. Randall R, Davies H, Patterson F, Farrell K. Selecting doctors for postgraduate training in paediatrics using a competency based assessment centre. *Archives of Disease in Childhood*. 2006;91(5):444-8.
158. Randall R, Stewart P, Farrell K, Patterson F. Using an assessment centre to select doctors for postgraduate training in obstetrics and gynaecology. *The Obstetrician & Gynaecologist*. 2006;8(4):257-62.
159. Roberts MJ, Gale TC, Sice PJA, Anderson IR. The relative reliability of actively participating and passively observing raters in a simulation-based assessment for selection to specialty training in anaesthesia. *Anaesthesia*. 2013;68(6):591-9.
160. Ackerman PL. A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*. 1996;22(2):227-57.
161. Ackerman PL, Heggestad ED. Intelligence, personality, and interests: evidence for overlapping traits. *Psychological Bulletin*. 1997;121(2):219-45.
162. O'Neill L, Vonsild MC, Wallstedt B, Dornan T. Admission criteria and diversity in medical school. *Medical Education*. 2013;47(6):557-61.
163. Work Psychology Group. SJT UKCAT Technical report 2013. April 2014.
164. Tiffin PA, Dowell JS, McLachlan JC. Widening access to UK medical education for under-represented socioeconomic groups: modelling the impact of the UKCAT in the 2009 cohort. *British Medical Journal*. 2012 2012-04-17 22:32:20.
165. Patterson F, Prescott-Clements L, Zibarras L, Edwards H, Kerrin M, Cousans F. Recruiting for values in healthcare: A preliminary review of the evidence. *Advances in Healthcare Sciences Education*. In press.
166. Julian ER. Validity of the Medical College Admission Test for Predicting Medical School Performance. *Academic Medicine*. 2005;80(10):910-7.
167. Evans P, Wen FK. Does the medical college admission test predict global academic performance in osteopathic medical school? *The Journal of the American Osteopathic Association*. 2007;107(4):157-62.
168. Cliffordson C, Askling B. Different Grounds for Admission: Its effects on recruitment and achievement in medical education. *Scandinavian Journal of Educational Research*. 2006;50(1):45-62.
169. Osman NY, Schonhardt-Bailey C, Walling JL, Katz JT, Alexander EK. Textual analysis of internal medicine residency personal statements: themes and gender differences. *Medical Education*. 2015;49(1):93-102.

- 1
2
3 170. Dore KL, Hanson M, Reiter HI, Blanchard M, Deeth K, Eva KW. Medical school admissions:
4 enhancing the reliability and validity of an autobiographical screening tool. *Academic medicine :
5 journal of the Association of American Medical Colleges*. 2006;81(10 Suppl):S70-3.
- 6 171. Chan D, Schmitt N. Video-based versus paper-and-pencil method of assessment in situational
7 judgment tests: subgroup differences in test performance and face validity perceptions. *The Journal of
8 Applied Psychology*. 1997;82(1):143-59.
- 9 172. Dore KL, Reiter HI, Eva KW, Krueger S, Scriven E, Siu E, et al. Extending the interview to
10 all medical school candidates--Computer-Based Multiple Sample Evaluation of Noncognitive Skills
11 (CMSENS). *Academic Medicine: Journal of the Association of American Medical Colleges*.
12 2009;84(10 Suppl):S9-12.
- 13 173. Lievens F, Patterson F. The validity and incremental validity of knowledge tests, low-fidelity
14 simulations, and high-fidelity simulations for predicting job performance in advanced-level high-
15 stakes selection. *The Journal of Applied Psychology*. 2011;96(5):927-40.
- 16 174. McDaniel MA, Morgeson FP, Finnegan EB, Campion MA, Braverman EP. Use of situational
17 judgment tests to predict job performance: a clarification of the literature. *The Journal of Applied
18 Psychology*. 2001;86(4):730-40.
- 19 175. Schubert S, Ortwein H, Dumitsch A, Schwantes U, Wilhelm O, Kiessling C. A situational
20 judgement test of professional behaviour: development and validation. *Medical Teacher*.
21 2008;30(5):528-33.
- 22 176. Bore M, Munro D, Kerridge I, Powis D. Selection of medical students according to their
23 moral orientation. *Medical Education*. 2005;39(3):266-75.
- 24 177. Dowell J, Lumsden MA, Powis D, Munro D, Bore M, Makubate B, et al. Predictive validity
25 of the personal qualities assessment for selection of medical students in Scotland. *Medical Teacher*.
26 2011;33(9):e485-8.
- 27 178. Lumsden MA, Bore M, Millar K, Jack R, Powis D. Assessment of personal qualities in
28 relation to admission to medical school. *Medical Education*. 2005;39(3):258-65.
- 29 179. Tsou KI, Lin CS, Cho SL, Powis D, Bore M, Munro D, et al. Using personal qualities
30 assessment to measure the moral orientation and personal qualities of medical students in a non-
31 Western culture. *Evaluation & the Health Professions*. 2013;36(2):174-90.
- 32 180. Dahlin M, Soderberg S, Holm U, Nilsson I, Farnebo LO. Comparison of communication
33 skills between medical students admitted after interviews or on academic merits. *BMC Medical
34 Education*. 2012;12:46.
- 35 181. Eva KW, Reiter HI, Rosenfeld J, Trinh K, Wood TJ, Norman GR. Association between a
36 medical school admission process using the multiple mini-interview and national licensing
37 examination scores. *The Journal of the American Medical Association*. 2012;308(21):2233-40.
- 38 182. Harris S, Owen C. Discerning quality: using the multiple mini-interview in student selection
39 for the Australian National University Medical School. *Medical Education*. 2007;41(3):234-41.
- 40 183. Lemay JF, Lockyer JM, Collin VT, Brownell AK. Assessment of non-cognitive traits through
41 the admissions multiple mini-interview. *Medical Education*. 2007;41(6):573-9.
- 42 184. Lambe P, Waters C, Bristow D. Do differentials in the support and advice available at UK
43 schools and colleges influence candidate performance in the medical school admissions interview? A
44 survey of direct school leaver applicants to a UK medical school. *Medical Teacher*. 2013;35(9):731-9.
- 45 185. Taylor CA, Green KE, Spruce A. Evaluation of the effect of socio-economic status on
46 performance in a Multiple Mini Interview for admission to medical school. *Medical Teacher*.
47 2015;37(1):59-63.
- 48 186. Kreiter CD, Solow C, Brennan RL, Yin P, Ferguson K, Huebner K. Examining the influence
49 of using same versus different questions on the reliability of the medical school preadmission
50 interview. *Teaching and Learning in Medicine*. 2006;18(1):4-8.
- 51 187. Kelley SR, Ray MA, Tsuei BJ. Trends concerning a common medical school interview
52 question, revisited. *Medical teacher*. 2007;29(5):e139-42.
- 53 188. Raghavan M, Martin BD, Burnett M, Aoki F, Christensen H, Mackalski B, et al. Multiple
54 mini-interview scores of medical school applicants with and without rural attributes. *Rural Remote
55 Health*. 2013;13:2362.
- 56
57
58
59
60

189. Till H, Myford C, Dowell J. Improving student selection using multiple mini-interviews with multifaceted Rasch modeling. *Academic Medicine: Journal of the Association of American Medical Colleges*. 2013;88(2):216-23.

190. Brownell K, Lockyer J, Collin T, Lemay JF. Introduction of the multiple mini interview into the admissions process at the University of Calgary: acceptability and feasibility. *Medical Teacher*. 2007;29(4):394-6.

Table 1. Literature search specification

Databases Searched
<ul style="list-style-type: none"> • EBSCO • EMBASE • Educational Resources Information Center (ERIC) • SCOPUS • Web of Knowledge (WoK)
Search terms and strategy
<p>“medical school” or “medical student” or “medical education” AND “selection” or “admission” or “criteria” or “test” or “interview” or “predictive” or “psychometric” or “personality” or “resume” or “cv” or “curriculum vitae” or “application form” or “biodata” or “reference” or “sjt” or “situational judgment test” or “situational judgement test” or “selection centre” or “selection center” or “assessment centre” or “assessment center” or “emotional intelligence” or “ei” or “aptitude test” or “validity” and “reliability” or “construct”.</p>

Table 2. Research questions and evidence quality categories

Typology of Research Questions	
(a) Type of evidence	<i>What type of evidence is available?</i>
(b) Effectiveness	<i>Does this work?</i>
	<i>Does doing this work better than doing that?</i>
(c) Procedural issues	<i>What are the implementation issues?</i>
	<i>What are the limitations of using the selection method?</i>
(d) Acceptability	<i>How widely used is the selection method?</i>
	<i>Will medical schools be willing to or want to use the selection method?</i>
	<i>Are users, providers, and other stakeholders satisfied with the selection method?</i>
(e) Cost effectiveness	<i>What is the financial impact of the selection method for the medical school and student</i>
Study Type	
	• Meta-analysis, systematic review
	• Non-systematic review
	• Longitudinal study
	• Cross-sectional quantitative study
	• Cross-sectional qualitative

Figure 1. Flowchart of study inclusion and exclusion

Potentially relevant articles identified by search strategy and title screened for relevance (n=1,407)

⇒ Clearly irrelevant articles excluded (n=1,079). Studies did not relate to selection methods in a relevant context.

⇓ Title and abstracts screened for evaluation regarding inclusion criteria (n=326)

⇒ Articles excluded based on contents of title and abstract (n=28)

⇓ Full texts screened for evaluation regarding exclusion criteria (n=298)

⇒ Articles excluded based on contents of full text (n=121)

⇓ **Studies included in review (n=179)**

².

² The results section provides a summary of the evidence from the literature. For a full list and description of all papers identified in the review, refer to Tables 3 and 4.

Table 3. Articles relating to each selection method and research question category

	Research Questions				
Selection Methods	(a) Effectiveness	(b) Procedural issues	(c) Acceptability	(d) effectiveness Cost	Total Articles for Each Selection Method
Aptitude Tests	36 articles (1-4, 10, 11, 13, 14, 16-21, 23-27, 29-37, 52, 55, 57, 66, 108, 115, 166, 167)	11 articles (22, 24, 38-46)	6 articles (12, 15, 24, 28, 42, 47)	0 articles	50 articles
Academic Records	25 articles (3, 6, 11, 14-19, 23, 27, 32, 36, 50-59, 61, 79)	5 articles (18, 49, 62, 63, 168)	2 articles (12, 47)	0 articles	31 articles
Personal Statements	9 articles (1, 13, 59, 60, 65, 66, 69, 72, 169)	6 articles (67, 68, 70-72, 170)	2 articles (64, 72)	0 articles	15 articles
References	5 articles (13, 60, 71, 73, 74)	1 article (75)	3 articles (13, 71, 75)	0 articles	6 articles
SJTs	16 articles (76-78, 80-85, 90, 92, 171-175)	6 articles (78, 82, 86-89)	9 articles (77, 82, 89-95)	1 article (82)	24 articles
Personality Assessment	20 articles (56, 60, 65, 73, 96-101, 103, 104, 110-112, 116, 176-179)	3 articles (97, 104, 110)	2 articles (110, 111)	1 article (98)	20 articles
Emotional Intelligence	6 articles (102, 105-109)	1 article (109)	0 articles	1 article (109)	6 articles
Interviews	53 articles (1, 3, 8, 13-15, 17, 19, 27, 29, 32, 64, 66, 69, 73, 113-130, 132-134, 136-141, 143, 149, 150, 152, 180-186)	16 articles (1, 44, 48, 71, 111, 120, 124, 125, 133, 143-145, 186-189)	11 articles (1, 128, 131, 141, 142, 146-149, 152, 190)	5 articles (135, 140, 151, 152, 190)	70 articles

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Selection Centres	4 articles (153-155, 159)	1 article (159)	6 articles (153-158)	1 article (155)	7 articles
Total Articles for Each Research Questions	134 articles	48 articles	39 articles	9 articles	-

For Review

Table 4. Summary of the relevant findings for each selection method

Aptitude Tests

Authors & Year	Reference number	Study type	Summary of relevant findings/conclusions
Kreiter CD, Axelson RD. (2013)	(1)	Non-systematic review	Despite mixed results in the research evidence, MCAT was concluded to be a robust indicator of who will perform well in the medical profession.
Husbands A, Dowell J. (2013)	(66)	Longitudinal	UKCAT different predictive validity across two cohorts at Dundee Medical School. For the 2010 cohort, UKCAT had no correlations with students' success in early years (OSCE and written examination). However in an earlier cohort (2009), UKCAT scores explained 6% of the variance in semester 1 and 2 written examination, and 7% of the variance in semester 1 OSCE in combination with an MMI.
Brannick M, Grichanik M, Nazian S, Wahi M, Goldin S. (2013)	(108)	Longitudinal	MCAT predicted all medical school outcomes better than any other predictor.
Tiffin PA, McLachlan JC, Webster L, Nicholson S. (2014)	(47)	Cross sectional, quantitative	Most of the sociodemographic factors that predict A level attainment also predict UKCAT performance. However, compared to A levels, males and those speaking English as a first language perform better on UKCAT. UKCAT scores may be more influenced by sex and less sensitive to school type compared to A levels.
Simpson PL, Scicluna HA, Jones PD, Cole AM, O'Sullivan AJ, Harris PG, Velan G, McNeil HP. (2014)	(32)	Longitudinal	The UMAT did not predict any performance outcomes at medical school.
Sartania N, McClure JD, Sweeting H,	(19)	Longitudinal	UKCAT has a modest predictive power for overall course performance at the University of Glasgow Medical School over and above that of school science achievements (UCAS score) or pre-admission interview score.

Browitt A. (2014)			
Puddey IB, Mercer A, Andrich D, Styles I. (2014)	(31)	Cross sectional, quantitative	Re-sitting the UMAT augments performance in each of its components together with the total UMAT percentile score. Whether this increase represents just an improvement in performance or an improvement in understanding of the variables and therefore competence needs to be further defined.
Puddey IB, Mercer A. (2014)	(14)	Longitudinal	Total GAMSAT score was a consistent independent predictor of academic performance as measured by the weighted average mark for the core units completed throughout a graduate entry medical programme. GAMSAT Section 3 (Reasoning in the biological and physical sciences) with Section 1 (Reasoning in the humanities and social sciences) and Section 2 (Written communication) also contributed either later or earlier in the course respectively.
Poole P, Shulruf B. (2013)	(30)	Longitudinal	The best predictor of a 'strong' interest in general practice was a low UMAT score of between 45 and 55 on all three UMAT sections. Yet the academic scores at entry of students with these UMAT scores were not lower than those of their classmates.
Moore EJ, Price DL, Van Abel KM, Carlson ML. (2014)	(24)	Longitudinal	A practical test for a residency programme showed good inter-rater reliability. Both the overall aptitude test scores and the subset attitudinal score showed reliability in predicting performance during residency training.
McManus IC, Dewberry C, Nicholson S, Dowell JS, Woolf K, Potts HWW. (2013)	(57)	Meta-analysis/ Longitudinal	Aptitude tests significantly predicted undergraduate and postgraduate performance, but much less well than academic indicators.
Laurence CO, Zajac IT, Lorimer M, Turnbull DA, Sumner KE. (2013)	(29)	Cross sectional, quantitative	Applicants who attended training courses on the UMAT by private organisations, used the online services of private organisations, or who familiarised themselves with the process were significantly more likely to receive an offer of an interview than those who did not undertake the activity. The odds of being offered an interview increased with each preparatory activity undertaken.
Husbands A,	(17)	Longitudinal	UKCAT appeared to predict performance better in the later years of medical

Mathieson A, Dowell J, Cleland J, MacKenzie R. (2014)			school compared to in the earlier years, and provided modest supportive evidence for the UKCAT's role in student selection in two UK medical schools.
Griffin B, Yeomans ND, Wilson IG. (2013)	(28)	Cross sectional, quantitative	Despite having higher academic grades on entry, students who had been coached for the UMAT had a lower GPA at medical school.
Edwards D, Friedman T, Pearce J. (2013)	(15)	Longitudinal	UMAT scores correlated with performance in the first two years of medical school, largely driven by the Logical Reasoning and Understanding People sections of the UMAT.
McManus IC, Dewberry C, Nicholson S, Dowell JS. (2013)	(18)	Longitudinal	Performance at UKCAT did correlate with first-year performance at medical school. The correlation was small but significant for secondary school leavers and was larger for mature entrants. The incremental validity of UKCAT after taking the current educational attainment used for selection into account was small but significant.
Adam J, Dowell J, Greatrix R. (2011)	(38)	Cross sectional, qualitative	There was considerable variation in how UK medical schools used UKCAT results to inform decision-making in medical student selection during 2006-2010.
Al-Rukban MO, Munshi FM, Abdulghani HM, Al-Hoqail I. (2010)	(36)	Longitudinal	An aptitude selection test was not predictive of medical students' GPA. Details of the aptitude test were not provided.
Albanese MA, Farrell P, Dottl S. (2005)	(39)	Longitudinal	Different statistical methods for determining MCAT cut-scores discriminated differently between candidates in relation to their performance the on medical licensing examination (USMLE Step 1).
Aldous CJ, Leeder SR, Price J, Sefton AE, Teubner JK. (1997)	(43)	Cross sectional, quantitative	Scores for GAMSAT varied significantly with candidate sex, age, highest degree level, and main subject in first degree. Mean scores were highest for men, younger candidates, honours graduates, and those with a physical sciences background.
Bell JF. (2005)	(21)	Non-systematic review	The BMAT predicted performance on university examinations, and the skills that are assessed by the BMAT are used by doctors in both practice and research.

Emery JL, Bell JF, Vidal Rodeiro CL. (2011)	(22)	Longitudinal	Despite some differences in applicants' BMAT performance (e.g. by school type and gender), BMAT scores predicted mean examination marks equitably for all background variables considered.
McManus IC, Ferguson E, Wakeford R, Powis D, James D. (2011)	(35)	Non-systematic review	Call for fuller presentation of data and more robust statistical analysis of the BMAT.
Callahan CA, Hojat M, Veloski J, Erdmann JB, Gonnella JS. (2010)	(10)	Longitudinal	MCAT had short and long-term predictive validity for medical school performance, attrition, scores on the medical licensing examinations, and ratings of clinical competence in the first year of residency. There was differential validity for men (higher) and women (lower).
Cleland JA, French FH, Johnston PW. (2011)	(42)	Cross sectional, quantitative & qualitative	Only 20% of first year medical students agreed that the UKCAT was useful in the selection procedure. Focus groups identified four themes related to views of the UKCAT: lack of face validity, concerns about fairness and cost, the use of data by medical schools, and influence of preparation.
Coates H. (2008)	(2)	Longitudinal	GAMSAT scores added value to the other data that are factored into selection decisions, and had a reasonable relationship with student marks in Year 1, affirming the valuable role of GAMSAT in medical school selection.
Donnon T, Paolucci EO, Violato C. (2007)	(25)	Meta-analytic, systematic review	The predictive validity of the MCAT ranged from small to medium for both medical school performance and medical board licensing examination measures. The medical profession is challenged to develop screening and selection criteria with improved validity that can supplement the MCAT as an important criterion for admission to medical schools.
Elam CL, Stratton TD, Scott KL, Wilson JF, Lieber A. (2002)	(12)	Cross sectional, quantitative	MCAT scores were one of the most influential factors in determining decisions made by medical school admissions committee members.
Evans P, Wen FK. (2007)	(167)	Longitudinal	The MCAT had limited predictive value in determining global academic performance in osteopathic medical school (GPAs and licensing examination scores).
Griffin B, Harding DW, Wilson IG,	(44)	Cross sectional, quantitative	Coaching had a small positive effect on the non-verbal reasoning component of the UMAT.

Yeomans ND. (2008)			
Halpenny D, Cadoo K, Halpenny M, Burke J, Torreggiani WC. (2010)	(4)	Cross sectional, quantitative	HPAT predicted medical school performance.
Hissbach J, Klusmann D, Hampe W. (2011)	(40)	Cross sectional, tool development	A multidimensional HAM-Nat test was expected to be a better selection tool than a uni-dimensional version of the test.
Julian ER. (2005)	(166)	Longitudinal	MCAT scores performed well as an indicator of academic preparation for medical school, independent of GPA scores.
Kreiter CD, Kreiter Y. (2007)	(52)	Meta-analytic, systematic review	MCAT had a positive predictive relationship with clinical skills. A validity generalization perspective supported the use of the MCAT for selection into medical school.
Lambe P, Waters C, Bristow D. (2012)	(45)	Cross sectional, quantitative	UKCAT performance was associated with differentials in access to support and advice, modes of preparation, type of school/college, level of achievement in mathematics, gender and age.
McManus IC, Ferguson E, Wakeford R, Powis D, James D. (2011)	(34)	Longitudinal	BMAT section 1 (science knowledge and applications) was predictive of medical school performance, while section 2 (aptitude and skills) was not.
McManus IC, Smithers E, Partridge P, Keeling A, Fleming PR. (2003)	(55)	Longitudinal	Aptitude tests had little predictive validity for subsequent medical careers.
Peskun C, Detsky A, Shandling M. (2007)	(13)	Longitudinal	MCAT was predictive of medical school performance.
Poole P, Shulruf B, Rudland J,	(16)	Longitudinal	UMAT had small significant incremental validity over GPA in predicting performance in medical school.

Wilkinson T. (2012)			
Trost G, Nauels HU, Klieme E. (1998)	(3)	Longitudinal	Highest pass rates in first medical examination were achieved by those selected on basis of school leaving certificate and aptitude test (Test for Medical Studies).
Wilkinson D, Zhang J, Byrne GJ, Luke H, Ozolins IZ, Parker MH, Peterson RF. (2008)	(27)	Longitudinal	GPA, interview and GAMSAT score were only modestly predictive of academic performance, and GAMSAT was the weakest predictor.
Wilkinson D, Zhang J, Parker M. (2011)	(33)	Longitudinal	UMAT had limited predictive validity for academic performance.
Wright SR, Bradley PM. (2010)	(20)	Longitudinal	UKCAT scores were predictive of year 1 and 2 examination performance at medical school.
Yates J, James D. (2010)	(26)	Longitudinal	The predictive validity of the UKCAT was low. Section scores may predict success in specific types of course assessment.
Zhao X, Oppler S, Dunleavy D, Kroopnick M. (2010)	(41)	Cross sectional, quantitative	There are multiple methods for using repeaters' MCAT scores to predict medical school performance. Average score may be a better approach than most recent, highest-within-administration, and highest-across-administration.
Albishri JA, Aly SM, Alnemaary Y. (2012)	(23)	Longitudinal	QUDRAAT was statistically predictive of GPA.
Dunleavy DM, Kroopnick MH, Dowd KW, Searcy CA, Zhao X. (2013)	(11)	Longitudinal	The combination of GPA and MCAT total scores performed well as a predictor of performance. Both GPA and MCAT total scores were strong predictors of academic performance in medical school through graduation, not just the first two years. These relationships generalized across medical schools.
Kraft HG, Lamina C, Kluckner T, Wild C, Prodinger WM.(2013)	(37)	Longitudinal	Students selected using an aptitude test were more able and better motivated to study medicine than those selected not using one.

Winegarden B, Glaser D, Schwartz A, Kelly C. (2012)	(46)	Longitudinal	MCAT's verbal reasoning component differed in predictive validity for English language and non-English language students.
---	------	--------------	---

Academic Records

Authors & Year	Reference number	Study type	Summary of relevant findings/conclusions
Puddey IB, Mercer A. (2014)	(14)	Longitudinal	GPA at entry was a consistent independent predictor of academic performance as measured by the weighted average mark for the core units completed throughout a graduate entry medical programme.
Husbands A, Mathieson A, Dowell J, Cleland J, MacKenzie R. (2014)	(17)	Longitudinal	UCAS form scores appeared to lack predictive validity, yielding no statistically significant positive associations with year 4 medical school performance.
Sartania N, McClure JD, Sweeting H, Browitt A. (2014)	(19)	Longitudinal	School science achievements (UCAS score) was less predictive than UKCAT of overall course performance at the University of Glasgow Medical School.
Simpson PL, Scicluna HA, Jones PD, Cole AM, O'Sullivan AJ, Harris PG, Velan G, McNeil HP. (2014)	(32)	Longitudinal	Academic achievement was the best predictor of overall and knowledge-based outcomes at medical school.
Tiffin PA, McLachlan JC, Webster L, Nicholson S. (2014)	(47)	Cross sectional, quantitative	Most of the sociodemographic factors that predicted A level attainment also predicted UKCAT performance.
Ferguson E, Semper H, Yates J, Fitzgerald JE,	(56)	Longitudinal	The effects of intelligence were limited to early learning, with intelligence predicting GCSE and pre-clinical knowledge but unrelated to the rest of the learning process, as the effect of A levels on clinical knowledge was

Skatova A, James D. (2014)			significantly different from its effect on clinical skills.
McManus IC, Dewberry C, Nicholson S, Dowell JS, Woolf K, Potts HWW. (2013)	(57)	Meta-analysis/ Longitudinal	A-levels were better predictors than GCSEs for undergraduate and postgraduate performance. Both showed construct-level predictive validities with undergraduate performance, and were slightly more predictive than aptitude tests.
McManus IC, Dewberry C, Nicholson S, Dowell JS. (2013)	(18)	Longitudinal	Educational attainment was clearly a strong predictor of medical school achievement, but it is currently limited by a large number of applicants getting top grades. AS-levels and GCSE results both had incremental value over A-levels (and a wider range of performance). AS-levels and GCSEs also have the practical advantage of being available at the time of selection, rather than merely being estimated grades for examinations yet to be taken.
Edwards D, Friedman T, Pearce J. (2013)	(15)	Longitudinal	School achievement generally had stronger correlations with medical school GPA than admissions interviews and UMAT scores.
Tektas OY, Fiessler C, Mayr A, Neuhuber W, Paulsen F. (2013)	(61)	Longitudinal	There was a weak association between high school examination grades and the grades achieved in the written and oral part of the first medical state examination.
Urlings-Strop LC, Stegers-Jager KM, Stijnen T, Themmen AP. (2013)	(59)	Longitudinal	Selection procedure applicants (academic and non-academic measures) had 4.4% lower dropout rate than lottery admitted applicants.
Luqman M. (2013)	(54)	Longitudinal	Pre-admission scores had a significant, moderate correlation with academic success in 1 st year examinations, which became weaker in professional examinations in higher classes.
Bhatti MA, Anwar M. (2012)	(50)	Longitudinal	Applicants who performed well in FSC (academic records) also performed well in entry tests for medical school and subsequent study, except year 1.

Al-Rukban MO, Munshi FM, Abdulghani HM, Al-Hoqail I. (2010)	(36)	Longitudinal	High school grades were not predictive of medical school GPA.
Elam CL, Stratton TD, Scott KL, Wilson JF, Lieber A. (2002)	(12)	Cross sectional, quantitative	Factors cited as influencing admission committee members' preliminary votes on applicants after initial screening, final votes after committee deliberation, and written comments regarding issues of concern influencing their votes cast (in declining order of frequency) included: MCAT scores, medical experience, comparison with other applicants, grades, letters of evaluation, interviews, individual attributes, residency status, service experience, expressed desire of committee members to discuss the applicant at the meeting, American Medical College Application Service personal statement, and diversity.
Kreiter CD, Kreiter Y. (2007)	(52)	Meta-analytic, systematic review	Relevant studies suggested that MCAT and undergraduate GPA have a positive predictive relationship with clinical skills.
McManus IC, Powis DA, Wakeford R, Ferguson E, James D, Richards P. (2005)	(62)	Non-systematic review	A levels, which used a more finely developed marking system at the top end (A+ and A++ grades, for example) had the greatest potential towards enabling enhanced selection by medical schools' admissions staff. Such grades would be maximally robust, in view of the testing time (and coursework) involved.
McManus IC, Smithers E, Partridge P, Keeling A, Fleming PR. (2003)	(55)	Longitudinal	Results of A level grades, which are particularly used for selection of students in the United Kingdom, had long-term predictive validity for undergraduate and postgraduate careers. In contrast, a test of ability or aptitude (AH5) was of little predictive validity for subsequent medical careers.
O'Neill L, Hartvigsen J, Wallstedt B, Korsholm L, Eika B. (2011)	(58)	Longitudinal	Students admitted on grades alone had a higher chance of dropping out than those admitted based on an admission test.
Poole P, Shulruf B,	(16)	Longitudinal	The ability of the general cognitive test UMAT to predict outcomes in major

Rudland J, Wilkinson T. (2012)			assessments within medical programmes was relatively minor in comparison with that of the admission GPA, but the UMAT score added a small amount of predictive power when it was used in combination with the GPA.
Trost G, Nauels HU, Klieme E. (1998)	(3)	Cross sectional, quantitative	Scores on the First Medical Examination were highest for students admitted on the basis of high school grades in combination with an aptitude test.
Wilkinson D, Zhang J, Byrne GJ, Luke H, Ozolins IZ, Parker MH, Peterson RF. (2008)	(27)	Longitudinal	GPA was more strongly associated with medical score performance than GAMSAT and interview scores.
Albishri JA, Aly SM, Alnemaary Y. (2012)	(23)	Longitudinal	High school grades were statistical predictive of GPA.
Dunleavy DM, Kroopnick MH, Dowd KW, Searcy CA, Zhao X. (2013)	(11)	Longitudinal	The combination of GPA and MCAT total scores performed well as a predictor of performance. Both GPA and MCAT total scores were strong predictors of academic performance in medical school through to graduation, not just the first two years. These relationships generalized across medical schools.
O'Flynn S, Fitzgerald T, Mills A. (2013)	(49)	Cross sectional, quantitative	Aptitude test (HPAT) scores remained stable while school leaving grades were increasing.
Cliffordson C, Askling B. (2006)	(168)	Longitudinal	Admissions based on upper secondary grades best promoted the goal of diversity.
Cohen-Schotanus J, Muijtjens AM, Reinders JJ, Agsteribbe J, van Rossum HJ, van der Vleuten CP. (2006)	(51)	Longitudinal	The GPA of school-leaving examinations was found to be related to study success, career development and scientific performance.
Ferguson E, James D, Madeley L. (2002)	(6)	Meta-analytic, systematic review	Previous academic performance was a good but not perfect predictor of achievement in medical training. It accounted for 23% of the variance in undergraduate training and 6% of the variance in postgraduate training.

Hänsel M, Klupp S, Graupner A, Dieter P, Koch T. (2010)	(79)	Longitudinal	School leaving GPA seemed to be the best predictor of success on the First National Medical Examination.
Lumb AB, Vail A. (2004)	(53)	Longitudinal	School-leaving grades were significant predictors of success in the OSCE.
McManus IC, Woolf K, Dacre J. (2008)	(63)	Longitudinal	UK A-level grades did not provide sufficient discriminatory power to enable selection of the most able students.

Personal Statements

Authors & Year	Reference number	Study type	Summary of relevant findings/conclusions
Kreiter CD, Axelson RD. (2013)	(1)	Non-systematic review	Faking responses and coaching may represent an insurmountable barrier to the use of personal statements in medical school admissions.
Husbands A, Dowell J. (2013)	(66)	Longitudinal	The UCAS personal statement form had no correlations with students' success in early years (OSCE and written examination) at Dundee Medical School across two separate cohorts.
Kumwenda B, Dowell J, Husbands A. (2013)	(72)	Cross sectional, quantitative	Most applicants believed that other applicants stretch the truth on their personal statement. Applicants with lower UKCAT scores (<2600) were more likely to believe that people stretch the truth on their personal statement and agree it acceptable for them to put fraudulent information in.
Osman NY, Schonhardt-Bailey C, Walling JL, Katz JT, Alexander EK. (2015)	(169)	Cross sectional, quantitative	Some themes were common to all applications. Male applicants were more likely to describe personal attributes and self-promote, while woman more frequently expressed the communicative and team-based aspects of doctoring.
Elam, CL, Johnson MM. (1997)	(69)	Cross sectional, quantitative	Applicants considered earlier in the selection process were more likely to gain admission, despite no difference in academic qualifications. Non-cognitive characteristics of applicants related to the medical school's mission predicted admission decisions.
Elam CL, Stratton TD, Scott KL,	(12)	Cross sectional, quantitative	The contents of medical school candidates' application forms were very unlikely to exert any significant influence on decisions made by admissions committees.

Wilson JF, Lieber A. (2002)			
Peskun C, Detsky A, Shandling M. (2007)	(13)	Longitudinal	Non-cognitive assessment (reviews of an autobiographical sketch, personal essay and reference letters by three evaluators) correlated significantly with ranking in internal medicine.
Dore KL, Hanson M, Reiter HI, Blanchard M, Deeth K, Eva KW. (2006)	(170)	Cross-sectional, quantitative	The reliability/validity of an autobiographical screening tool was affected by scoring method (evaluating each candidate in turn versus evaluating all candidates for each question in turn). Evaluating candidates for each question in turn resulted in greater reliability & prediction.
Ferguson E, James D, O'Hehir F, Sanders A, McManus IC. (2003)	(60)	Longitudinal	Personal statements were predictive of clinical aspects of training.
Ferguson E, Sanders A, O'Hehir F, James D. (2000)	(65)	Longitudinal	Scores on personal statements were not predictive of subsequent success at a medical school overall.
Hanson MD, Dore KL, Reiter HI, Eva KW. (2007)	(70)	Cross sectional, quantitative	Scores given to autobiographical submissions were higher for submissions completed off-site compared to on-site. The two sets of ratings were not correlated with one another. On-site ratings increased with time allowed for completion, but the reliability of the measure was not affected by increased time.
Oosterveld P, ten Cate O. (2004)	(64)	Cross sectional, quantitative	Autobiographical submissions had low reliability compared to other common selection instruments. The contents of written testimonies including autobiographical submissions were not likely to reflect the genuine nature of medical school candidates as successfully as selection methods like interviews or observations.
Parry J, Mathers J, Stevens A, Parsons A, Lilford R, Spurgeon P, et al. (2006)	(71)	Cross sectional, qualitative	There was inconsistency between UK medical schools in terms of how data from autobiographical submissions were used. Some medical schools formally used the information in making selection decisions, while others ignored this information due to concerns that it may unfairly bias selection decisions.

White J, Brownell K, Lemay JF, Lockyer JM. (2012)	(67)	Cross sectional, qualitative	There was a tension between “genuine” and “expected” responses in admissions essays, which applicants face when choosing how to answer questions in the admissions process.
White JS, Lemay JF, Brownell K, Lockyer J. (2011)	(68)	Cross sectional, qualitative	There was a disconnect between the approach of applicants (to show themselves in order to be selected as individuals) and the stated intent of the process (to select applicants based on objective criteria).

References

Authors & Year	Reference number	Study type	Summary of relevant findings/conclusions
Peskun C, Detsky A, Shandling M. (2007)	(13)	Longitudinal	Non-cognitive assessment (including reference letters) correlated with ranking in internal medicine, but not with ranking in family medicine.
Benbassat J, Baumal R. (2007)	(73)	Non-systematic review	No significant association between letters of recommendation and medical school outcome measures were found across two published peer-reviewed studies.
Ferguson E, James D, O'Hehir F, Sanders A, McManus IC. (2003)	(60)	Longitudinal	Information in teachers' references was not a consistent predictor of performance during a medical degree. Teacher's references were concluded to have no practical use as a source of information about applicants.
Parry J, Mathers J, Stevens A, Parsons A, Lilford R, Spurgeon P, et al. (2006)	(71)	Cross sectional, qualitative	Medical schools differed in relation to how they used the information presented in referees' reports. Some schools made use of this information, while others ignored it because of concerns over bias.
Poole PJ, Moriarty HJ, Wearn AM, Wilkinson TJ, Weller JM. (2009)	(74)	Non-systematic review	Personal references have been shown to be of no predictive value.
Stedman JM, Hatch	(75)	Cross-sectional,	Reference writers applied positive and negative attributions homogenously

JP, Schoenfeld LS. (2009)		quantitative	across applicants, thus rendering applicant differentiation on this basis impossible.
---------------------------	--	--------------	---

Situational Judgement Tests

Authors & Year	Reference number	Type of Study	Summary of relevant findings/conclusions
Rostom H, Watson R, Leaver L. (2013)	(76)	Non-systematic review	SJTs may be preferable to the previous 'white space' questions used in UK foundation programme selection, but coaching was a confounding variable that was identified as requiring further consideration and research.
Patterson F, Ashworth V, Kerrin M, O'Neill P. (2013)	(85)	Non-systematic review	SJTs can be designed to be less susceptible to coaching by tailoring their content and the response formats used and instructions given.
Libbrecht N, Lievens F, Carette B, Cote S. (2014)	(84)	Longitudinal	Emotional intelligence measured by an SJT showed incremental validity over conscientiousness and cognitive ability for predicting interpersonal academic performance.
Cullen, MJ, Sackett, PR, Lievens, F. (2006)	(88)	Cross sectional, quantitative	Training was effective in raising scores on a class-based test (CSQ) but not on an SJT.
Christian, MS, Edwards BD, Bradley, JC. (2010)	(78)	Meta-analytic, systematic review	Constructs measured by SJTs were not always specified, but SJTs can be developed to assess specific constructs, most often leadership and interpersonal skills. Video-based SJTs were more strongly correlated with performance than pencil and paper SJTs.
Clevenger J, Pereira GM, Wiechmann D, Schmitt N, Harvey VS. (2001)	(93)	Cross sectional, quantitative	An SJT was found to be a valid predictor of performance in three samples, and an incremental predictor over job knowledge, cognitive ability, job experience and conscientiousness in two samples.
O'Connell MS, Hartman NS, McDaniel MA, Grubb WL, Lawrence A. (2007)	(94)	Cross sectional, quantitative	An SJT had incremental validity over cognitive ability and personality for predicting task and contextual performance.
Ahmed H,	(92)	Longitudinal	An SJT was predictive of performance in workplace-based selection centre

Rhydderch M, Matthews P. (2012)			simulations, and was a better predictor of performance than a clinical problem-solving test. However, the greatest degree of predictive validity was provided by combining results from the SJT and clinical problem-solving test. The SJT was concluded to play a valuable role in shortlisting.
Cabrera MAM, Nguyen NT. (2001)	(77)	Meta-analytic, systematic review	SJTs had significant predictive and criterion related validity for job performance across numerous industries including medicine.
Chan D, Schmitt N. (1997)	(171)	Cross-sectional, quantitative	The validity of SJTs may differ for video-based SJTs versus paper-and-pencil SJTs.
Chan D, Schmitt N. (2002)	(89)	Non-systematic review	Across a number of professions, SJTs had incremental validity over the prediction provided jointly by cognitively ability, personality assessment and job experience.
Dore KL, Reiter HI, Eva KW, Krueger S, Scriven E, Siu E, et al. (2009)	(172)	Cross-sectional, quantitative	A video-based SJT had strong psychometric properties including MMI correlation.
Koczwara A, Patterson F, Zibarras L, Kerrin M, Irish B, Wilkinson M. (2012)	(90)	Longitudinal	An SJT was a better predictor of selection centre performance than a cognitive ability test and a clinical problem solving test. Candidate reactions to the SJT were more positive than reactions to the cognitive ability test.
Lievens F, Buyse T, Sackett PR. (2005)	(81)	Longitudinal	A video-based SJT was differentially valid for predicting overall GPA for different curricula. The SJT showed incremental validity over cognitively oriented measures for curricula that included interpersonal courses, but not for other curricula. The SJT became more valid through the years.
Lievens F, Patterson F. (2011)	(173)	Longitudinal	SJT performance was significantly associated with job performance among family doctors. The SJT had incremental validity over a knowledge test.
Lievens F, Peeters H, Schollaert E. (2008)	(82)	Non-systematic review	Across a number of professions including medicine, SJTs showed criterion-related and incremental validity over cognitive ability and personality tests. SJTs also had a less adverse impact towards minorities. Applicant reactions to SJTs were positive, and SJTs enabled testing of large applicant groups at once.

Lievens F, Sackett PR. (2007)	(86)	Multiple cohort study	Numerous different approaches existed to development of alternate forms of an individual SJT. These could affect alternate-form consistency, mean score changes, and criterion-related validity. It was recommended that consideration should be given to the alternate-form development process.
McDaniel MA, Hartman NS, Whetzel DL, Grubb WL. (2007)	(87)	Meta-analytic, systematic review	Across a range of professions including medicine, SJTs had incremental validity over cognitive ability, personality assessment and a composite of the two.
McDaniel MA, Morgeson FP, Finnegan EB, Champion MA, Braverman EP. (2001)	(174)	Meta-analytic, systematic review	Across a range of professions including medicine, SJTs had useful levels of criterion-related and construct validity.
Patterson F, Baron H, Carr V, Plint S, Lane P. (2009)	(95)	Longitudinal	An SJT was a valid shortlisting method, and had higher predictive validity than application form questions and a clinical problem-solving test.
Patterson F, Carr V, Zibarras L, Burr B, Berkin L, Plint S, et al. (2009)	(83)	Longitudinal	An SJT showed good reliability in a sample of Core Medical Trainee applicants. SJT was a good predictor of interview performance, and offered incremental validity over the existing shortlisting process.
Plint S, Patterson F. (2010)	(91)	Non-systematic review	Postgraduate speciality trainee candidate reactions towards SJT were favourable.
Schubert S, Ortwein H, Dumitsch A, Schwantes U, Wilhelm O, Kiessling C. (2008)	(175)	Psychometric test development study	Practical difficulties existed in obtaining expert consensus regarding best answers for use in SJTs in medical selection. Different methods for determining expert consensus (ranking and rating scales) may result in different outcomes.
Lievens F. (2013)	(80)	Longitudinal	Interpersonal skills assessment carried out using SJTs had significant added value over cognitive ability tests for predicting interpersonal GPA throughout the curriculum, doctor performance, and performance on an OSCE and in a

			case-based interview.
--	--	--	-----------------------

Personality & Emotional Intelligence

Authors & Year	Reference number	Study type	Summary of relevant findings/conclusions
Hojat M, Erdmann JB, Gonnella JS. (2013)	(97)	Non-systematic review	There was evidence for all Big Five personality traits as predictors of various performance outcomes at medical school, however conscientiousness and patient care were the two most conceptually relevant and valid predictors of success at medical school and beyond. Personality may be measured avoiding the confound of socially desirable responding. Explanations for the modest validity of personality measures in medical education research may include: multidimensionality of personality, construct dissimilarity, changes in predictor-criterion matching, proximal and distal criterion measures, restriction of range, nonlinear relationships, multicollinearity, variation in methods of assessment.
Tsou KI, Lin CS, Cho SL, Powis D, Bore M, Munro D et al. (2013)	(179)	Cross sectional, quantitative	Both tests of non-cognitive traits in the Personal Qualities Assessment (Mojac, moral orientation and NACE, involved/ detached personality test) demonstrated acceptable psychometric properties. However, the predictive validity of PQA requires further investigation.
Ferguson E, Semper H, Yates J, Fitzgerald JE, Skatova A, James D. (2014)	(56)	Longitudinal/ meta analysis	Conscientiousness was a positive predictor of pre-clinical knowledge and a negative predictor of clinical knowledge. Emotional stability was negatively associated with clinical skills. Selection models need to consider the different type of learning outcome when being developed, as simply selecting on a trait – on the assumption that it will always confer benefits – needs re-evaluating.
Brannick MT, Grichanik M, Nazian SJ, Mahi M, Goldin SB (2013)	(108)	Longitudinal	The WLEIS (self-report measure) was not significantly correlated with any of the measures of success in medical school. The MSCEIT (ability measure) showed significant correlations where students with higher EI scores tended to have better grades for both clinical years three and four. Self-assessments for EI were suggested to be unlikely to prove useful in admissions
Cherry MG, Fletcher I, O’Sullivan H, Dornan T. (2014).	(109)	Systematic review	There was insufficient evidence to support the use of EI as a selection criterion. The authors suggested that it may be appropriate to include EI in the curriculum rather than as selection criteria, as EI can be learned.

Lin DT, Kannappan A, Lau JN. (2013)	(106)	Cross sectional, quantitative	Applicant EI correlated poorly with academic parameters and was not accurately assessed by faculty interviews. The authors suggested that methods that better capture EI should be incorporated into the residency selection process.
Edwards JC, Elam CL, Wagoner NE. (2001)	(103)	Non-systematic review	Proposed a measure for medical school admission; emphasised the importance of qualitative variables such as compassion, altruism, respect and integrity.
Elam CL, Studts JL, Johnson MMS. (1997)	(116)	Longitudinal	Contents of interview reports, including applicants' attributes, correlated positively with medical school performance.
Powis DA, Rolfe I. (1998)	(112)	Non-systematic review	Selection procedure based on desirable personal qualities may have important social benefits as it can enable greater diversity of academically well-qualified applicants to enter medical school.
Lievens F, Coetsier P, De Fruyt F, Maeseneer J. (2002)	(98)	Cross sectional, quantitative	Extraversion and agreeableness (dimensions defining interpersonal dynamic) may be beneficial for communication and collaboration skills in doctors. Conscientiousness affects examination results. The authors suggested personality assessment may be useful a tool for student counselling and guidance.
Bore M, Munro D, Powis D. (2009)	(104)	Non-systematic review	Personality had relatively low but consistent and significant predictive validity coefficients in relation to work performance.
Benbassat J, Baumal R. (2007)	(73)	Non-systematic review	A moderate correlation existed between personality measures and performance at medical school.
Ferguson E, James D, O'Hehir F, Sanders A, McManus IC. (2003)	(60)	Longitudinal	Conscientiousness was consistently a better predictor of performance in medical school than teachers' references and A-level grades. Conscientiousness was positively related to preclinical performance but negatively related to clinical grades.
Ferguson E, Sanders A, O'Hehir F, James D. (2000)	(65)	Longitudinal	Conscientiousness was significantly related to success in medical training, and demonstrated incremental validity over previous academic performance.
Bore M, Munro D, Kerridge I, Powis	(176)	Longitudinal	Moral orientation scores were found to be correlated with a number of personality measures, providing evidence of construct validity. Moral

D. (2005)			orientation significantly predicted moral decisions made in response to hypothetical dilemmas.
Carr SE. (2009)	(105)	Cross-sectional, quantitative	EI was not significantly associated with measures of cognition and skill in medical students.
Carrothers RM, Gregory SW, Jr., Gallagher TJ. (2000)	(102)	Cross-sectional, quantitative	An EI measure was developed that demonstrated the ability to measure attributes that indicate desirable personal and interpersonal skills in medical school applicants.
Chan-Ob T, Boonyanaruthee V. (1999)	(101)	Longitudinal	Three of the scales of the California Personality Inventory (dominance, flexibility and socialization) were positively related to GPA among medical students, while two of the scales (sociability and sense of well-being) were negatively related to GPA.
Dowell J, Lumsden MA, Powis D, Munro D, Bore M, Makubate B, et al. (2011)	(177)	Longitudinal	No significant correlations were found between separate elements of the Personal Qualities Assessment and performance in medical school. "Non-extreme" character types on involved-detached and libertarian-communitarian scales ranked higher on OSCEs on average.
Haight SJ, Chibnall JT, Schindler DL, Slavin SJ. (2012)	(96)	Cross-sectional, quantitative	Personality variables were associated with non-cognitive indicators of performance at medical school (clinical evaluations, humanism nominations). Conscientiousness predicted clinical skills, extraversion predicted clinical skills reflecting interpersonal behaviour, and empathy predicted motivation.
Jerant A, Griffin E, Rainwater J, Henderson M, Sousa F, Bertakis KD, et al. (2012)	(111)	Cross-sectional, quantitative	Extraversion was associated with MMI scores. Extraversion and agreeableness were associated with medical school acceptance offers.
Knights JA, Kennedy BJ. (2006)	(99)	Cross-sectional, quantitative	Current medical students may have negative personality traits, which are not identified by selection methods such as interviews.
Knights JA, Kennedy BJ. (2007)	(100)	Cross-sectional, quantitative	Certain negative personality traits may be associated with lower performance in relation to aspects of academic performance in medical school.
Leddy JJ, Moineau G, Puddester D,	(107)	Longitudinal	EI measures were not significantly correlated with other traditional measures used to determine medical school admission. Therefore, EI may not currently

Wood TJ, Humphrey-Murto S. (2011)			be assessed in traditional admissions processes.
Lievens F, Ones DS, Dilchert S. (2009)	(110)	Longitudinal	Extraversion, openness and conscientiousness scores showed increases in operational validity for predicting GPA. Although there may not be any advantages to being open and extraverted for early academic performance, these traits gained importance for later academic performance when applied practice increasingly plays a part in the curriculum.
Lumsden MA, Bore M, Millar K, Jack R, Powis D. (2005)	(178)	Cross-sectional, quantitative	The authors concluded that the incorporation of the Personal Qualities Assessment tool into medical school selection systems may have positive implications for widening access and training doctors who are more representative of the community at large.

Interviews

Authors & Year	Reference number	Study type	Summary of relevant findings/conclusions
Kreiter CD, Axelson RD. (2013)	(1)	Non-systematic review	The traditional interview should not be an influential component in selection, and the use of an interview score to make the final decision on a candidate's success may violate their expectation of fair and valid assessment practice. MMIs have demonstrated the feasibility of generating a reliable non-academic assessment from an interview-like procedure and have displayed acceptable reliabilities and promising validity evidence.
Husbands A, Dowell J. (2013)	(66)	Longitudinal	An MMI was the most consistent predictor of success in early years (OSCE and written examination) at Dundee Medical School across two separate cohorts.
Husbands A, Mathieson A, Dowell J, Cleland J, Mackenzie R. (2014)	(17)	Longitudinal	Traditional interview scores appeared to lack predictive validity, and yielded no statistically significant positive associations with year 4 medical school performance.
Laurence CO, Zajac IT, Lorimer M, Turnall DA, Sumner KE. (2013)	(29)	Cross sectional, quantitative	Two preparatory activities for the oral assessment interview (refining and clearing a personal resume, and learning the course structure) were significantly associated with being offered a place in medical school.

Simpson PL, Scicluna HA, Jones PD, Cole AM, O'Sullivan AJ, Harris PG, Velan G, McNeil HP. (2014)	(32)	Longitudinal	The communication skills dimension of the structured admission interview was a significant predictor of clinical skills, but predictive and unique variance values were small.
Puddey IB, Mercer A. (2014)	(14)	Longitudinal	Interview score only weakly predicted performance later in a graduate entry medical programme, and mainly in clinically-based units
Edwards D, Friedman T, Pearce J. (2013)	(15)	Longitudinal	Poorer performance on the interview is associated with greater performance in medical school GPA.
Sartania N, McClure JD, Sweeting H, Browitt A. (2014)	(19)	Longitudinal	Pre-admission interview score was less predictive of overall course performance than UKCAT at the University of Glasgow Medical School.
Casey M, Wilkinson D, Fitzgerald J, Eley D, Connor J. (2014)	(127)	Longitudinal	No difference was found between students who had completed a medical school admission interview and those who had not, in clinical communication scores including (active listening, warmth and rapport, eye contact, appropriate questions, empathic responding, cue identification). English as a first language and gender were more significant predictors of communication skills scores than a selection interview.
Lambe P, Waters C, Bristow D. (2013)	(184)	Cross sectional, quantitative	Differences in medical school applicants' interview score were partially explained by differences in access to advice and support at the school or college a respondent attended during the application stage for a place at medical school.
Campagna-Vaillancourt M, Manoukian J, Razack S, Nguyen LH. (2014)	(134)	Cross sectional, quantitative	The majority of applicants (>80%) felt that the MMI helped them to present their strengths and was free of any gender, cultural, or age bias. Most assessors (>85%) agreed the MMI evaluated a valid range of competencies, and that it tested more aspects of an applicant than did traditional interviews. Both applicants and assessors (>70%) agreed that the MMI was a fair process, and both preferred it over the traditional interview. Overall, interrater reliability of

			the MMI was good.
Taylor CA, Green KE, Spruce A. (2015)	(185)	Cross sectional, quantitative	Applicants from nonselective, non-fee-paying schools and/or from areas with traditionally lower HE participation did slightly better in the MMI, but this difference was not statistically significant.
Eva KW, Macala C. (2004)	(128)	Cross sectional, quantitative	No difference in interviewer rating was found between unstructured and structured MMI stations, although behavioural indicator stations showed better capacity to consistently discriminate between applicants relative to the other forms of interview (situational judgement and unstructured). Candidates considered the unstructured stations to be more challenging and more anxiety-provoking than either of the structured stations.
Eva KW, Reiter HI, Rosenfeld J, Trinh K, Wood TJ, Norman GR. (2012)	(181)	Longitudinal	Compared with students who were rejected by an admission process that used MMI assessment, students who were accepted scored higher on Canadian national licensing examinations.
Eva KW, Reiter HI, Rosenfeld J, Norman GR. (2004)	(136)	Cross sectional, quantitative	While further validity testing is required, the MMI appeared better able to predict preclerkship performance relative to traditional tools designed to assess the noncognitive qualities of applicants.
Hissbach JC, Sehner S, Harendza S, Hampe W. (2014)	(135)	Cross sectional, quantitative	The authors suggested that it would be more cost-efficient to raise the number of stations instead of raising the number of raters within stations to increase the reliability of the HAM-Int. Different stations assess different attributes. Rater bias was observed as having a greater effect on applicant scores than systematic differences in candidate performance.
Hopson LR, Burkhardt JC, Stansfield RB, Vohra T, Turner-Lawrence D, Losman ED. (2014)	(141)	Cross sectional, quantitative	Although the MMI alone was viewed less favourably than a traditional interview, participants were receptive to a mixed-methods interview (MMI and traditional interview). The MMI correlated with performance on an emergency medicine clerkship.
Pau A, Jeevaratnam K, Chen YS, Fall AA, Khoo C, Nadarajah VD.	(142)	Meta-analysis/ Systematic review	Findings suggested that MIMIs are feasible in terms of efficient utilisation of time, costs and human resources when compared to the panel interview; generally acceptable to both interviewees and interviewers; generally reliable and predictive of future performance in certain aspects of medical council

(2013)			examinations.
Tiller D, O'Mara D, Rothnie I, Dunn S, Lee L, Roberts C. (2013)	(152)	Cross sectional, quantitative	Although two-thirds of candidates would prefer an in-person interview, the cost and time savings for candidates were substantial when conducting an iMMI online via Skype. The authors suggested that the iMMI process could readily be applied by other medical schools or by professional colleges for postgraduate training.
Sebok SS, Luu K, Klinger DA. (2014)	(150)	Cross sectional, quantitative	Increasing the number of stations in an MMI would lead to a greater reduction in error variance compared to adding more raters per station.
Elam CL, Johnson MM. (1997)	(69)	Cross sectional, quantitative	Interview ratings, along with academic qualifications, applicant demographics, and educational development accounted for 39% of the variance in final admission status of applicants who received interviews.
Griffin B, Harding DW, Wilson IG, Yeomans ND. (2008)	(44)	Cross sectional, quantitative	Coaching did not assist and may even hinder performance on an MMI. Nevertheless, as practice on similar MMI tasks did improve scores, the authors suggested that tasks should be rotated each year. Further research is required on the predictive validity of the UMAT, given that coaching appeared to have a small positive effect on the non-verbal reasoning component of the test.
Peskun C, Detsky A, Shandling M. (2007)	(13)	Longitudinal	An admissions interview correlated with residency ranking in family medicine (but not internal medicine).
Rahbar MH, Vellani C, Sajan F, Zaidi AA, Akbarali L. (2001)	(119)	Longitudinal	Interview ratings were associated with physiology scores but not with anatomy, biochemistry and community health sciences examinations held after five trimesters.
Trost G, Nauels HU, Klieme E. (1998)	(3)	Longitudinal	Students admitted on the basis of an interview had lower pass rates on First Medical Examination than those admitted on the basis of aptitude test/GPA.
Wilkinson D, Zhang J, Byrne GJ, Luke H, Ozolins IZ, Parker MH, et al. (2008)	(27)	Longitudinal	The school's selection criteria only modestly predicted performance in 1st/4th year examinations. GPA was most strongly associated, followed by interview and GAMSAT.

1 2 3 4 5 6 7 8 9 10	Basco WT, Jr., Lancaster CJ, Gilbert GE, Carey ME, Blue AV. (2008)	(123)	Longitudinal	The admission ranking and interview process did not predict clinical performance or patients' satisfaction on an OSCE.
11 12 13 14 15 16 17 18 19 20	Benbassat J, Baumal R. (2007)	(73)	Non-systematic review	Most studies had not detected any predictive validity of admission interviews for medical school grades, success on licensing examinations, and intern performance ratings. Some authors had reported that high scores on admission interviews predicted achievement of honours, the Dean's letters of recommendation on graduation, OSCE scores, ranking for admission for Family Medicine residency, performance on tests of diagnostic reasoning and communication skills, while low scores predicted withdrawal rates from medical schools. However, the extent to which these outcomes reflect non-cognitive traits, cognitive aptitude, or both is uncertain.
21 22	Oosterveld P, ten Cate O. (2004)	(64)	Cross sectional, quantitative	An interview had lower validity than a study sample assessment procedure, but higher validity than application form.
23 24 25 26 27 28	Parry J, Mathers J, Stevens A, Parsons A, Lilford R, Spurgeon P, et al. (2006)	(71)	Cross sectional, qualitative	Some schools did not interview; some shortlisted for interview only on predicted academic performance while those that shortlist on a wider range of non-academic criteria use various techniques and tools to do so. Once students were shortlisted, interviews varied in terms of length, panel composition, structure, content, and scoring methods.
29 30 31 32 33	Jerant A, Griffin E, Rainwater J, Henderson M, Sousa F, Bertakis KD, et al. (2012)	(111)	Cross sectional, quantitative	Extraversion was associated with MMI performance, whereas both extraversion and agreeableness were associated with acceptance offers. Adoption of the MMI may affect diversity in medical student personalities, with potential implications for students' professional growth, specialty distribution, and patient care.
34 35 36 37 38	Axelson R, Kreiter C, Ferguson K, Solow C, Huebner K. (2010)	(143)	Cross sectional, quantitative	Unstructured interviews had higher reliability than structured interviews. Combining scores from both yielded the highest reliability scores.
39 40 41	Donnon T, Oddone- Paolucci E, Violato	(114)	Longitudinal	A semi-structured interview based on clearly defined and scored medical judgment vignettes that focus on the assessment of medical students' non-

42
43
44
45
46
47
48
49

C. (2009)			cognitive attributes is promising for student's selection into medical school.
Dowell J, Lynch B, Till H, Kumwenda B, Husbands A. (2012)	(148)	Cross sectional, quantitative	The strategy of generating institutional support for MMIs through staged introduction proved effective. The MMI in Dundee Medical School was shown to be feasible and displayed sound psychometric properties. Student assessors appeared to perform at least as well as staff.
Eva KW, Reiter HI, Rosenfeld J, Norman GR. (2004)	(130)	Cross sectional, quantitative	The reliability (internal consistency) of the MMI was observed to be 0.65. The variance component attributable to candidate-station interaction was greater than that attributable to candidate. So context specificity may reduce the validity of interviews. Both applicants and examiners were positive about the experience and the potential for this protocol.
Harris S, Owen C. (2007)	(182)	Cross sectional, quantitative	MMIs proved to be an efficient process by which to interview candidates and to determine suitability. Retained and rejected candidates had significantly different total scores and mean scores for each station. Ten independent observations contributed to each decision, without significant interviewer or logistic burden. Candidates reported high levels of satisfaction with the interview process.
Kreiter C, Yin P, Solow C, Brennan R. (2004)	(125)	Cross sectional, quantitative	Interview scores derived from standardised interviews were found to display low to moderate levels of reliability. The authors concluded that interview scores do not appear to possess the level of precision found with other measures commonly used to facilitate admissions decisions.
McManus IC, Richards P, Winder BC. (1999)	(147)	Cross sectional, quantitative	Applicants preferred medical schools that conduct interviews.
O'Brien A, Harvey J, Shannon M, Lewis K, Valencia O. (2011)	(132)	Cross sectional, quantitative	The authors concluded that MMIs were reliable, feasible, and acceptable to both applicants and interviewers. Longitudinal research is needed to establish the validity of MMIs.
Patrick LE, Altmaier EM, Kuperman S, Ugolini K. (2001)	(118)	Cross sectional, quantitative	Structured interviews had moderate-to-low correlations with GPA/MCAT/application forms, suggesting that the interview provided information about candidate credentials not obtained from other sources, and accounted for a substantial proportion of the variance in admission status. Findings supported the considerable time and resources required to develop a

			structured interview for medical student admissions. Longitudinal research is needed to assess the validity and utility of the method.
Quintero AJ, Segal LS, King TS, Black KP. (2009)	(145)	Cross sectional, quantitative	Interviewers rated applicants with personalities similar to theirs most favourably.
Roberts C, Walton M, Rothnie I, Crossley J, Lyon P, Kumar K, et al. (2008)	(133)	Cross sectional, quantitative	An MMI was moderately reliable. The largest source of error was interviewer subjectivity, suggesting interviewer training is beneficial. An MMI needs to be sufficiently long for precise comparison for ranking purposes. The concurrent validity of the MMI was supported by a small positive correlation with GAMSAT section scores.
Till H, Myford C, Dowell J. (2013)	(189)	Cross sectional, quantitative	Quality control monitoring is essential to ensure fairness when ranking candidates according to scores obtained in the MMI.
Ann Courneya C, Wright K, Frinton V, Mak E, Schulzer M, Pachev G. (2005)	(113)	Cross sectional, quantitative	An MMI had good inter-panel reliability, high consistency within and between interview panels, and uniformly positive questionnaire responses about panel interviews versus individual interviews.
Basco WT, Jr., Gilbert GE, Chessman AW, Blue AV. (2000)	(121)	Cross sectional, quantitative	The admission ranking and interview process did not predict clinical performance or patients' satisfaction on an OSCE.
Brownell K, Lockyer J, Collin T, Lemay JF. (2007)	(190)	Cross sectional, qualitative	The MMI allowed assessors to interview applicants in one weekend, with fewer interviewers and less time required per interviewer compared to our previous interview process. More than 90% of both the applicants and interviewers found the process to be very acceptable.
Dodson M, Crotty B, Prideaux D, Carne R, Ward A, de Leeuw E. (2009)	(151)	Cross sectional, quantitative	Reducing the duration of MMI stations from 8 to 5 minutes conserved resources with minimal effect on applicant ranking and test reliability.
Donnon T, Paolucci EO. (2008)	(115)	Cross sectional, quantitative	The Medical Judgment Vignettes interview provided a reliable approach to the assessment of candidates' non-cognitive attributes for medical school. The MMI had high inter-rater reliability, attributed to greater objectivity and clearly

			defined scoring. There was a relatively high generalisability coefficient for only three stations, but future research should explore reliability and validity of vignettes in larger samples.
Dore KL, Kreuger S, Ladhani M, Rolfson D, Kurtz D, Kulasegaram K, et al. (2010)	(129)	Cross sectional, quantitative	The MMI provided a reliable way to assess residency candidates that is acceptable to both candidates and assessors across a variety of programs.
Eva KW, Reiter HI, Trinh K, Wasi P, Rosenfeld J, Norman GR. (2009)	(137)	Longitudinal	The correlation between performance on the MMI and number of stations passed on an objective structured clinical examination-based licensing examination was $r = 0.43$ ($P < 0.05$) in a postgraduate sample and $r = 0.35$ ($P < 0.05$) in an under-graduate sample of subjects who sat the MMI 5 years prior to sitting the licensing examination.
Griffin BN, Wilson IG. (2010)	(48)	Cross sectional, quantitative	Rating leniency was associated with personality and sex of interviewers, but the effect was small. Random allocation of interviewers, similar proportions of male and female interviewers across applicant interview groups, use of the MMI format, and skills-based interviewer training were all suggested to be likely to reduce the effect of variance between interviewers.
Hofmeister M, Lockyer J, Crutcher R. (2008)	(131)	Cross sectional, quantitative	Standardized residency selection interviews could be adapted to measure professionalism potential characteristics important to family medicine in ways that are acceptable to IMG applicants and interviewers.
Hofmeister M, Lockyer J, Crutcher R. (2009)	(138)	Longitudinal	There was evidence that the MMI offers a reliable and valid assessment of professionalism in IMG doctors applying for Canadian family medicine residencies and that this clinically situated MMI assessed facets of competency other than those assessed by the OSCE.
Humphrey S, Dowson S, Wall D, Diwakar V, Goodyear HM. (2008)	(149)	Cross sectional, quantitative	Both candidates and interviewers agreed that the MMI format was reliable, fair and asked appropriate, easy-to-understand questions. In high-stakes interviews such as for specialty training in Modernising Medical Careers programmes, it is vital that all concerned have confidence in the selection process.
Kleshinski J, Shriner C, Khuder	(117)	Cross sectional, qualitative	Professionalism scenarios can be a worthwhile tool for use in the admissions process. The interview process should encourage participation from faculty who

SA. (2008)			value this as an important component in the evaluation of an applicant. Future research should investigate determinants of faculty perception of the role of assessing professionalism in the interview process.
Kreiter CD, Solow C, Brennan RL, Yin P, Ferguson K, Huebner K. (2006)	(186)	Cross sectional, quantitative	Because there are a number of drawbacks to using the same questions for all applicants (i.e., security and validity) and little advantage in terms of increased reliability, the semi-structured question format should be considered when conducting the MSPI. The authors suggested a method of implementing a semi-structured interview is to present each applicant with a set of questions randomly drawn from a pool of interview questions.
Kumar K, Roberts C, Rothnie I, du Fresne C, Walton M. (2009)	(144)	Cross sectional, qualitative	Differing expectations regarding the MMI should be remediated through targeted interviewer training that aims to facilitate a shared understanding of what is being assessed and identify areas of potential interviewer bias in interacting with candidates and across scenarios. Concerns about benchmarking should be addressed by providing interviewers with marking practice on sample candidates of varying degrees of capability across different types of stations. Also propose that the MMI would benefit from the inclusion of a station that assesses the candidate's commitment to a career in medicine.
Lemay JF, Lockyer JM, Collin VT, Brownell AK. (2007)	(183)	Cross sectional, quantitative	The MMI was able to assess different non-cognitive attributes. The MMI offered a fairer and more defensible assessment of applicants to medical school than the traditional interview.
Prideaux D, Roberts C, Eva K, Centeno A, McCrorie P, McManus C, et al. (2009)	(8)	Non-systematic review	Despite the ubiquity of interviews (face-to-face contact with a single interviewer or a panel with varying degrees of structure), there were very few studies defining its psychometric properties. Those that do exist do not indicate that the interview is a robust selection measure.
Razack S, Faremo S, Drolet F, Snell L, Wiseman J, Pickering J. (2009)	(146)	Cross sectional, quantitative	The MMI was rated more highly than the traditional interview on fairness, imposition of stress and effectiveness as a measurement tool. Compared with the traditional interview, applicants also felt the MMI: (i) allowed them to be competitive; (ii) was enjoyable, and (iii) was often a favourite part of their interview experience. Evaluators' responses were in agreement with applicants'

			responses, albeit that overall they expressed more caution about the MMI.
Reiter HI, Eva KW, Rosenfeld J, Norman GR. (2007)	(139)	Longitudinal	The MMI complemented pre-admission cognitive measures to predict performance outcomes during clerkship and on the Canadian national licensing examination.
Rosenfeld JM, Reiter HI, Trinh K, Eva KW. (2008)	(140)	Non-systematic review	The MMI was more reliable and had better predictive power than the traditional panel interviews. The MMI requires greater preparatory efforts and a larger number of rooms to carry out the interviews relative to panel-based interviews, but that these cost disadvantages can be offset by the MMI requiring fewer person-hours of effort.
Basco WT, Lancaster C, Carey ME, Gilbert GE, Blue AV. (2004)	(122)	Longitudinal	The admission ranking and interview process did not predict clinical performance or patients' satisfaction on an OSCE.
Elam CL, Studts JL, Johnson MS. (1997)	(116)	Longitudinal	Data from interview reports were associated with GPA and performance on clerkship examinations.
Fan AP, Tsai TC, Su TP, Kosik RO, Morisky DE, Chen CH, et al. (2010)	(124)	Longitudinal	Students admitted through the interview route had a 3.20-point higher first-year medical school GPA. Those students who were admitted via interview did not have significantly different personality traits than those admitted through the traditional route. Results call into question the ability of an admissions interview to select for non-cognitive character traits.
Kelley SR, Ray MA, Tsuei BJ. (2007)	(187)	Cross sectional, quantitative & qualitative	Asking prospective medical students during their medical school interview what type of medicine they wish to practice may not yield valid responses.
Streyffeler L, Altmaier EM, Kuperman S, Patrick LE. (2009)	(126)	Longitudinal	Results did not support the predictive validity of an interview-based measure above other cognitive and non-cognitive admissions variables more easily gathered. However, in some domains, interview-based variables did incrementally predict medical school performance.
VanSusteren TJ, Suter E, Romrell LJ, Lanier L, Hatch RL. (1999)	(120)	Cross sectional, quantitative	Inter-rater agreement was found to be high, indicating that the trained raters were able to judge applicants' characteristics, independently of the quantitative data. Interviewers' scores and the MCAT Verbal accounted for a substantial amount of the variance in the selection decision. The authors concluded that

			medical school selection interviews can carry a significant and important weight in the selection decision. Interviews should be structured and conducted independently of the applicants' academic records.
Dahlin M, Soderberg S, Holm U, Nilsson I, Farnebo LO. (2012)	(180)	Cross sectional, quantitative	Students admitted through interview had better communication skills than those admitted through academic merit.
Raghavan M, Martin BD, Burnett M, Aoki F, Christensen H, Mackalski B, et al. (2013)	(188)	Cross sectional, quantitative	Despite graduates from rural and urban high schools having comparable GPA, there exists a rural–urban divide in MMI scores that could exacerbate the under-representation of rural students in medical schools.

Selection Centres

Authors & Year	Reference number	Study type	Summary of relevant findings/conclusions
Roberts MJ, Gale TCE, Sice, PJA, Anderson IR. (2013)	(156)	Cross sectional, quantitative	Large differences were found in scores given to candidates and acceptability of inter-rater reliability between trained assessors and participating healthcare staff. However it may be viable to use other healthcare staff rather than trained assessors for some but not all SC stations.
Gafni N, Moshinsky A, Eisenberg O, Zeigler D, Ziv A. (2012)	(153)	Cross sectional, quantitative	An SC was developed and tested for use in medical student selection. The SC had desirable psychometric properties and provisional evidence was presented on the validity of the selection method.
Gale TC, Roberts MJ, Sice PJ, Langton JA, Patterson FC, Carr AS, et al. (2010)	(156)	Longitudinal	Candidates and assessors demonstrated strong approval of the SC method. Inter-rater reliability was acceptable, and internal consistency of the method was high. The predictive validity of the SC method was good in relation to workplace performance during the first year of appointment in anaesthesia specialty training.
Randall R, Davies H, Patterson F,	(157)	Cross sectional, quantitative	Candidates perceived the SC to be a fair selection method. The SC method represents a feasible selection approach to postgraduate speciality training and

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Farrell K. (2006)			provides greater breadth and depth of information about candidates than does a structured interview.
Randall R, Stewart P, Farrell K, Patterson F. (2006)	(158)	Cross sectional, quantitative	Same as above.
ten Cate O, Smal K. (2002)		Cross sectional, quantitative	There was provisional evidence that the SC method has satisfactory internal consistency and inter-rater reliability. There was no evidence on predictive validity.
Ziv A, Rubin O, Moshinsky A, Gafni N, Kotler M, Dagan Y, et al. (2008)	(155)	Cross sectional, quantitative	There was evidence that the SC method had desirable internal consistency, inter-rater reliability, and test-retest correlation. The correlation between SC scores and cognitive test scores approached zero, reflecting the value of the SC in a selection process. The SC had high content and face validity.

For Review

Table 5. An interpretation of the wider literature relating to various selection methods

Selection Method	Implications of the Evidence
<i>Shortlisting Methods</i>	
Aptitude Tests	Evidence is mixed on the fairness and effectiveness of aptitude tests, specifically regarding predictive validity.
Academic Attainment	There is a high level of consensus regarding predictive validity, but concerns that the discriminatory power is diminishing as increasing numbers of students get top A Level grades.
Personal Statements	Candidate acceptability is high, but susceptibility to coaching is also high.
References	Use of references remains widespread despite little research supporting validity or reliability.
Situational Judgment Tests (SJTs)	Improved validity over other selection tools (IQ & personality tests), and can be mapped to organisational values. Whilst SJTs can be relatively costly to design, SJTs are machine-markable & can be delivered on-line, producing cost savings in high volume selection.
Personality Assessment and Emotional Intelligence	Where there is a high risk of susceptibility to faking and/or coaching, personality assessment is best used to drive more focused questioning at interviews (rather than a stand-alone instrument without verification). There is a dearth of long-term data for both EI and personality assessment. Research into EI is sparse, but initial evidence suggests that it may present a valuable tool in future medical selection.
<i>Final Stage Selection Methods</i>	
Traditional Interviews	Across most evaluation criteria, traditional interviews perform poorly.
Structured Interviews (e.g. competency-based and situational)	When interviews are structured and based on a thorough role analysis, with standardised questions with trained interviewers, and appropriate scoring they can be reliable and valid. Candidates prefer interviews to other methods although they are relatively resource intensive. (<i>Effective method for VBR</i>)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Multiple-mini interviews (MMIs),	MMIs are relatively expensive to design & implement, but may offer improved reliability and validity over one-to-one interviews. There remain some issues around construct validity of MMIs
Selection Centres (SCs) using work samples, e. g. group exercises, written/in-tray task, presentations, interactive exercises	SCs are relatively expensive to design & implement. Further evidence is needed of the predictive validity of SCs in undergraduate medical selection.

For Review