

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

Fiona Patterson,¹ Alec Knight,² Jon Dowell,³ Sandra Nicholson,⁴ Fran Cousans² & Jennifer Cleland⁵

CONTEXT Selection methods used by medical schools should reliably identify whether candidates are likely to be successful in medical training and ultimately become competent clinicians. However, there is little consensus regarding methods that reliably evaluate non-academic attributes, and longitudinal studies examining predictors of success after qualification are insufficient. This systematic review synthesises the extant research evidence on the relative strengths of various selection methods. We offer a research agenda and identify key considerations to inform policy and practice in the next 50 years.

METHODS A formalised literature search was conducted for studies published between 1997 and 2015. A total of 194 articles met the inclusion criteria and were appraised in relation to:

(vii) interviews and multiple mini-interviews (MMIs), and (viii) selection centres (SCs). The evidence relating to each method was reviewed against four evaluation criteria: effectiveness (reliability and validity); procedural issues; acceptability, and cost-effectiveness.

CONCLUSIONS Evidence shows clearly that academic records, MMIs, aptitude tests, SJTs and SCs are more effective selection methods and are generally fairer than traditional interviews, references and personal statements. However, achievement in different selection methods may differentially predict performance at the various stages of medical education and clinical practice. Research into selection has been over-reliant on cross-sectional study designs and has tended to focus on reliability estimates rather than validity as an indicator of quality. A comprehensive

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

CONTEXT Selection methods used by medical schools should reliably identify whether candidates are likely to be successful in medical training and ultimately become competent clinicians. However, there is little consensus regarding methods that reliably evaluate non-academic attributes, and longitudinal studies examining predictors of success after qualification are insufficient. This systematic review synthesises the extant research evidence on the relative strengths of various selection methods. We offer a research agenda and identify key considerations to inform policy and practice in the next 50 years.

METHODS A formalised literature search was conducted for studies published between 1997 and 2015. A total of 194 articles met the inclusion criteria and were appraised in relation to: (i) selection method used; (ii) research question(s) addressed, and (iii) type of study design.

RESULTS Eight selection methods were identified: (i) aptitude tests; (ii) academic records; (iii) personal statements; (iv) references; (v) situational judgement tests (SJTs); (vi) personality and emotional intelligence assessments; (vii) interviews and multiple mini-interviews (MMIs), and (viii) selection centres (SCs).

The evidence relating to each method was reviewed against four evaluation criteria: effectiveness (reliability and validity); procedural issues; acceptability, and cost-effectiveness.

CONCLUSIONS Evidence shows clearly that academic records, MMIs, aptitude tests, SJTs and SCs are more effective selection methods and are generally fairer than traditional interviews, references and personal statements. However, achievement in different selection methods may differentially predict performance at the various stages of medical education and clinical practice. Research into selection has been over-reliant on cross-sectional study designs and has tended to focus on reliability estimates rather than validity as an indicator of quality. A comprehensive framework of outcome criteria should be developed to allow researchers to interpret empirical evidence and compare selection methods fairly. This review highlights gaps in evidence for the combination of selection tools that is most effective and the weighting to be given to each tool.

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 and economic responsibility for medical education and training to produce competent clinicians in view of the high-stakes nature of the profession, and its relation to the health and well-being of individuals and societies, and financial cost. Indeed, selection for medical education internationally is frequently driven by political considerations and the preferences of key stakeholders.¹ Such influences may result in resistance against any move away from 'traditional' measures despite compelling evidence to do so, often to the detriment of evidence-based selection practices. However, Kreiter and Axelson's² 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 0.20), whereas evidence-based selection is comparatively far more powerful, with well-designed selection tools achieving performance gains exceeding one standard deviation. This finding is consistent with the historical progression of admissions strategies over the last 50 years, which have gradually moved away from subjective measures such as personal statements and references, towards more evidence-based models of selection. A remaining central concern refers to the determination of which of the different selection methods can reliably identify those who will be successful in medical training and ultimately become competent clinicians.

Over the last 50 years, selection for medicine has typically involved several different methods used in combination. Prior academic attainment has generally been, and continues to be, the primary basis for selection and is usually assessed at an initial screening stage.³ Academic indicators are typically used as the basis for initial shortlisting decisions in combination with personal statements, references or aptitude tests, which are usually followed by an interview at the final stage to support selection decisions. However, there are several concerns about this approach.

Firstly, 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 training and 6% in postgraduate performance.⁴ It is argued that academic ability is a necessary but not sufficient condition to ensure that a trainee becomes a competent clinician, and that there is a need over the next 50 years to increase the focus on the non-academic attributes important for success in clinical practice, as well as to develop and evaluate tools which assess such attributes.

Secondly, although academic achievement is consistently shown to be a good predictor of performance in medical school,⁵ historically substantially less attention has been paid to researching methods that reliably evaluate important non-academic personal attributes, interests and motivational qualities. It cannot be assumed that those with high academic ability alone can be turned into competent physicians via medical training, as other specific skills and qualities may need to be present from the start.⁶

Thirdly, there has been a dearth of longitudinal cohort studies examining the predictors of success after qualification. Specifically, there is a gap in research with respect to the long-term follow-up of trainees that links performance on different selection methods with subsequent performance in clinical practice.

Medical school admissions processes and selection for specialty training attract strong public interest and often criticism regarding fairness.^{7–9} There is a pressing need to review the research evidence on how best to combine selection methods and design valid selection systems to guide selectors in future with relation to, for example, the relative weightings to assign to academic and non-academic indicators in order to make valid selection decisions.

In order to explore these issues, we report here the results of a new systematic search and review of the research literature, examining studies in both undergraduate and postgraduate medical education. Specifically, we present the existing data on the relative strength of the research evidence for the quality of each of the various methods, as well as findings that may shape a future research agenda and inform future practice.

METHODS

Data sources

We conducted a formal literature search using the criteria specified in Table S1 (online). Our results were limited to English-language studies published between January 1997 and the end of April 2015. 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, FC and FP discussed these papers until all reviewers agreed on 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 undergraduate or postgraduate medical education (but not specialty training). We also included relevant systematic and meta-analytic reviews and non-systematic critical reviews, but excluded general opinion pieces, commentaries and letters. After the inclusion criteria had been applied, duplicate papers were removed, leaving the remaining articles to be retrieved for full-text review. Three authors (AK, FC and 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: (i) selection method type (e.g. interview, selection centre, etc.); (ii) research question addressed (e.g. cost-effectiveness, acceptability, etc. [see Muir and Grey 1996, cited in¹⁰]), and (iii) type of study design (e.g. meta-analysis, 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 the research questions for different selection methods. To generate a list of the different selection methods, AK and FC listed the selection method(s) assessed in each paper meeting the inclusion criteria, and asked an independent researcher to check the papers against the list for errors.

The research questions and evidence quality categories are displayed in Table 1.

Table 1 Research questions and categories of evidence quality

Typology of research questions	
1 Type of evidence	What type of evidence is available?
2 Effectiveness	Does this work? Does this approach work better than other approaches?
3 Procedural issues	What are the implementation issues? What are the limitations of using the selection method?
4 Acceptability	How widely used is the selection method? Will medical schools be willing to or want to use the selection method? Are users, providers and other stakeholders satisfied with the selection method?
5 Cost-effectiveness	What is the financial impact of the selection method for the medical school and student?
Study type	
Meta-analysis, systematic review	
Non-systematic review	
Longitudinal study	
Cross-sectional quantitative study	
Cross-sectional qualitative study	
Cross-sectional mixed-methods study	
Multiple cohort study	

In relation to the different research questions under investigation, we removed Muir and Grey's (1996)¹⁰ 'salience' and 'safety' categories as they were not relevant to our context. We also combined the 'acceptability' and 'appropriateness' categories, and refocused the 'procedural issues' category to more appropriately reflect the considerations given to implementing selection tools in medical education. Therefore, we examined each study in relation to four research questions concerning, respectively: effectiveness; procedural issues; acceptability, and cost-effectiveness. This approach was intended to address the assumption implicit in much previous research that predictive validity is the most important measure of the effectiveness of a selection method; we acknowledge that the success of a selection tool may be determined by a range of additional factors, including its accessibility, ease of implementation and the extent to which it is viewed as acceptable by key stakeholders. Finally, in relation to study quality, we categorised papers into seven general study types, including systematic and non-systematic reviews, longitudinal studies, quantitative, qualitative and mixed-methods. Studies reported within meta-analyses and systematic reviews were not assessed individually; rather, these are reported as the overall findings of each meta-analysis or systematic review.

RESULTS

The literature search produced 3152 hits across all databases including duplicates (EBSCO, $n = 732$; EMBASE, $n = 501$; ERIC, $n = 49$; Scopus, $n = 50$; Web of Knowledge, $n = 1820$). The titles and abstracts of the 3152 search results were scanned to remove obviously irrelevant articles and duplicates ($n = 2716$), leaving 436 articles for review. These abstracts were screened according to the eligibility criteria and a further 31 articles were removed (Fig. 1). All decisions were made by two researchers (AK and FC), but any uncertainties were discussed with another member of the research team (FP). Copies of the 405 articles were obtained and examined. Review of the full texts removed a further 211 articles. A total of 194 articles met the inclusion criteria for the present review (Appendix S1, online).

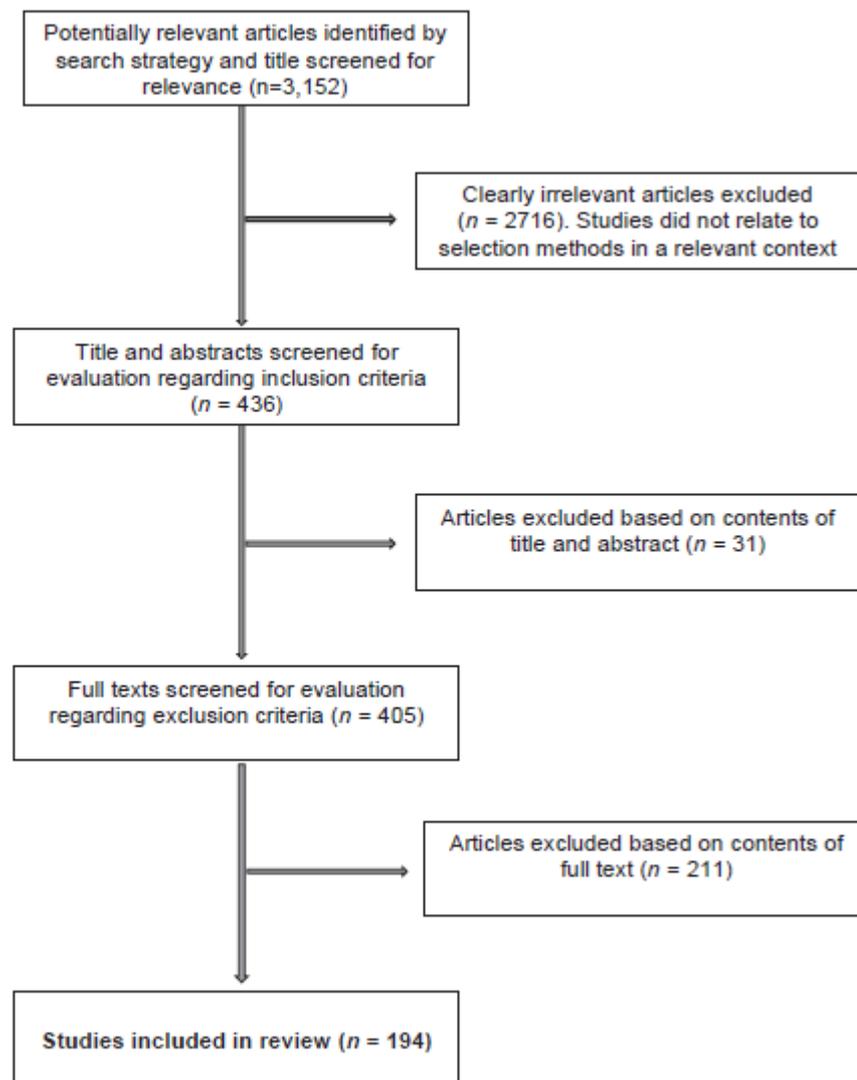


Figure 1 Flowchart of study inclusion and exclusion

The 194 studies were sorted into eight categories of different selection methods. Table S2 (online) shows the number of papers returned in relation to each selection method and research question, respectively. Studies investigating multiple selection methods or research categories were assigned to multiple categories, as required. A summary of the relevant review findings is presented online in Table S3. We acknowledge that there is a range of quality in the studies presented, irrespective of study type; however it is beyond the scope of this review to provide a detailed account of the quality of each study. Therefore, Table S3 is intended to provide a brief overview of the research evidence,

rather than a comprehensive description of each study. 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 S2 and S3 (online). We provide a more detailed overview of our synthesis of the research evidence below.

Aptitude tests

Type of evidence

Fifty-five studies were reviewed. Of these, three were systematic reviews or meta-analyses, three were non-systematic reviews, 34 were longitudinal (one was a meta-analysis), and 15 were cross-sectional (three mixed methods, one tool development, one qualitative, 10 quantitative).

Effectiveness

There is conflicting evidence on the predictive validity of aptitude tests in medical student selection. Some researchers have presented evidence to support the reliability and criterion-related, incremental or predictive validity for aptitude tests including the Medical College Admission Test (MCAT),^{11–14} the Graduate Australian Medical School Admissions Test (GAMSAT),^{15,16} the Undergraduate Medicine and Health Sciences Admissions Test (UMAT),^{17,18} the Health Professions Admissions Test (HPAT),⁵ the UK Clinical Aptitude Test (UKCAT),^{19–22} the Biomedical Admissions Test (BMAT)^{23,24} and Qudraat (the Saudi Arabian national aptitude examination).²⁵ Other researchers are sceptical of the effectiveness of the MCAT,²⁶ UKCAT,²⁷ GAMSAT,²⁸ UMAT,^{29–34} BMAT^{35,36} and an unspecified aptitude test.³⁷ However, some evidence suggests that students selected using an aptitude test may be more able and better motivated to study medicine than those selected using a process not including an aptitude test.³⁸ Finally, one paper³⁵ reported that Section 2 (science knowledge and applications) of the BMAT was predictive of medical school performance, whereas Section 1 (aptitude and skills) was not.

Procedural issues

Research suggests that variations in the ways in which aptitude tests are used in medical student selection may affect their reliability or validity.^{39–42} This is notable as medical schools vary in how they use aptitude tests to inform selection decisions and in the statistical methods used for determining cut scores and predicting subsequent performance. One article⁴¹ reported that the dimensionality of an aptitude test affected its effectiveness as a selection tool and that a scale composed of three subject-specific dimensions (biology, physics and chemistry) had better psychometric properties than a unidimensional model, even though the subject-specific scales were highly correlated and were used to calculate a global score.

Acceptability

One study¹³ reported that aptitude test scores were one of the most influential factors determining decisions made by medical school admissions committee members. However, other studies^{43,44} have reported that few Year 1 medical students agree that aptitude tests are a useful or suitable part of the selection procedure to medical school. Conversely, Stevens et al.⁴⁵ found that 76% of students agreed that the HPAT-Ireland is a fair test overall, but that Section 3 (non-verbal reasoning) appeared less acceptable and relevant than other sections.

Cost-effectiveness

No papers addressing the cost-effectiveness of aptitude tests were reviewed.

Summary

Mixed evidence exists among researchers on the usefulness of aptitude tests in medical student selection and findings largely depend on the specific aptitude test studied; hence commenting on the generality of findings is problematic. For example, some studies support the predictive validity of

aptitude tests, but other research suggests that some specific aptitude tests lack predictive validity. Mixed evidence also exists on the fairness of aptitude tests, with some research suggesting that certain groups score more highly on aptitude tests than other groups, whereas other research suggests that this is not the case. For example, there is varied evidence on the equity of aptitude tests for different groups of medical school applicants (e.g. according to sex, age, language status and socio-economic status).^{11,15,20,24,46–50} Other evidence suggests that aptitude tests are equitable with respect to candidate background, are affected relatively little by candidate coaching, and remain stable over time,^{20,24,44,50–52} with the possible exception of the UMAT.³⁰ It is therefore important to evaluate each aptitude test in its own right in order to draw conclusions on the quality of the tool.

Academic records

Type of evidence

Thirty-six studies assessing academic records were identified. Twenty-seven of these were longitudinal (one was a meta-analysis), two were meta-analyses, one was a non-systematic review, four were cross-sectional and quantitative, and two were cross-sectional and used mixed methods.

Effectiveness

Research evidence is generally highly concordant and supports the predictive validity of academic records in medical student selection.^{7,16,17,25,28,33,53–62} McManus et al.⁶³ describe how prior educational attainment forms the academic backbone of selection, progression through medical school and beyond. Another paper describes a small but significant incremental validity gain achieved by using candidates' educational achievement alongside aptitude tests compared with the use of traditional academic indicators alone.²⁰ International evidence also suggests that candidates admitted on the basis of their academic records had lower levels of dropout than those who were not.^{64,65} Incremental validity may be provided through the addition of an appropriate aptitude test.^{4,21,66} A minority of studies^{19,37,67} reported that academic records were not predictive of medical school performance.

Procedural issues

Some authors have argued that academic records may be unstable or may lack sufficient power to make fine distinctions between candidates.^{51,68,69} For example, McManus et al.^{68,69} posited that the current grading system for A-levels in the UK does not offer sufficient discriminatory power to enable the selection of the most able students.

Acceptability

Evidence was mixed on the acceptability of using academic records in medical student selection. This is illustrated by some authors citing academic records as an important factor that can influence selection decisions,¹³ and a range of opinion in how suitable students consider prior academic attainment to be for selection into medical school.^{44,45}

Cost-effectiveness

None of the papers reviewed addressed the cost-effectiveness of academic records.

Summary

There is a high level of consensus among researchers that academic records provide useful information to inform medical student selection. Research generally suggests that prior academic attainment has predictive power, meaning that those with stronger academic records are more likely to succeed in medical school. However, there is concern that the discriminatory power of prior academic attainment may be diminishing as increasing numbers of medical school applicants have top grades. There is also a lack of long-term follow-up data to provide evidence that medical school

applicants with higher grades go on to become better physicians. Moreover, Milburn⁸ notes that over-reliance on A-level results in the UK may create a distorted social intake to universities, and recruiting medical students solely on the basis of academic attainment may neglect important non-academic factors required for success in medical school and beyond. Further research is required to gauge the extent to which this is an international problem, and the potential value of contextual data in 'balancing' the requirement of academic attainment with access to medical education for individuals from non-traditional backgrounds.

Personal statements

Type of evidence

A total of 17 studies were reviewed, four of which were longitudinal. Twelve of the remaining studies were cross-sectional (four qualitative, seven quantitative, one mixed methods), and one was a non-systematic review.

Effectiveness

Evidence on the predictive validity of personal statements is varied. Although some evidence has been found for the predictive validity of personal statements for medical school dropout rates,⁶⁵ performance on internal medicine¹⁴ and clinical aspects of training,⁶⁶ several others have reported that personal statements have low reliability compared with other commonly used selection instruments⁷⁰ and are not predictive of subsequent success at medical school.^{2,71–73} Some authors suggest, however, that personal statements may have some value for making applicants aware of the characteristics of the medical degree they are applying to, which may help them to make a more informed decision to apply.⁷³

Procedural issues

Evidence suggests that a number of procedural factors affect the reliability and validity of personal statements. Medical school candidates may use personal statements to present themselves in ways they believe are attractive to admission committees, which may not necessarily be accurate.^{74,75} Hence, the information captured by personal statements is likely to be both partial and subjective in nature. Factors that may affect the effectiveness of the selection method include the earliness of submission in relation to a deadline,⁷⁶ marking method, and on-site versus off-site completion.⁷⁷ Finally, one article highlighted the fact that personal statements are used differentially by different UK medical schools.⁷⁸ Some medical schools use the information formally in making selection decisions, whereas others ignore this information out of concern that it may unfairly bias selection decisions.

Acceptability

Research has highlighted potential sources of data contamination in personal statements, including candidates' prior expectations, the length of time spent completing submissions, and input to submissions from third parties. Other research^{14,74} has commented on the political validity and stakeholder satisfaction of personal statements in medical student selection. Whereas Stevens et al.⁴⁵ found that approximately 60% of students thought that personal statements were suitable to use for admission to medical school, Elam et al.¹³ reported that the contents of medical school candidates' application forms are very unlikely to exert any significant influence on decisions made by admissions committees. White et al.⁷⁴ also argued that medical school candidates present themselves in ways that they believe are expected of candidates, rather than in ways that are genuine reflections of themselves. Likewise, Kumwenda et al.⁷⁹ found that most medical school applicants believed that others stretched the truth in their personal statements, and a proportion of applicants believed it was unlikely that statements were checked for accuracy.

Cost-effectiveness

No papers addressing the cost-effectiveness of personal statements were reviewed.

Summary

Evidence on the effectiveness of personal statements in medical student selection is mixed at best. Little evidence exists to support the predictive validity of personal statements, and a large volume of research evidence suggests that the selection method lacks reliability and validity. Personal statements remain widely used in medical school selection worldwide, despite concerns that the effectiveness of the selection method is influenced by numerous extraneous factors. The content of personal statements may also unfairly cloud the judgement of individuals making selection decisions.

References

Type of evidence

A total of nine articles were reviewed. Two were non-systematic reviews, three were longitudinal studies, and the remaining four were cross-sectional (one qualitative, one quantitative and two mixed methods).

Effectiveness

Studies examining the effectiveness of references did not usually include a direct empirical test of predictive validity,^{13,66,78,80,81} although there was some direct evidence^{66,82} that this selection method did not consistently predict performance at medical school. For example, Ferguson et al.⁶⁶ found that the information in teachers' references did not consistently predict medical school performance, and Poole et al.⁸¹ claimed that personal references have no predictive value. Overall, there was clear consensus among researchers that referees' reports were of limited use in predicting performance at medical school.

Procedural issues

One study⁸³ found that the content of the reports made it impossible for admissions committees to differentiate between applicants on the basis of the data they contained. Therefore, the authors⁸³ concluded that the utility of referees' reports in medical student selection was questionable at best.

Acceptability

Direct assessments of the acceptability of references were critical of the inclusion of referees' reports in medical student selection, and remarked that the information they contain may unduly bias admissions committees. One study commented that referees' reports remain widespread in medical student selection.⁷⁸ Medical students' perceptions of the acceptability of references were found to be varied.^{44,45}

Cost-effectiveness

Only one study commenting on the cost-effectiveness of references was found. De Zee et al.⁸² noted that letters of recommendation are expensive in terms of the time of admissions committee members, who must read and interpret each letter.

Summary

There is a good level of consensus that references are neither a reliable nor a valid tool for selecting candidates for medical school. Despite these findings, references remain a common feature of medical school selection worldwide. To this extent, the inclusion of references in medical school admission processes may be unhelpful and may use valuable resources that could be directed more usefully to selection methods with evidentially based reliability and validity.

Situational judgement tests

Type of evidence

A total of 25 studies focusing on situational judgement tests (SJTs) were reviewed. Of these, eight were longitudinal, six were cross-sectional quantitative studies, four were systematic reviews, and five were non-systematic reviews. Of the remaining two studies, one referred to the development of a test, and the other was a multiple cohort study.

Effectiveness

Despite some concern about susceptibility to coaching, 84–86 recent research has found that coaching and revising for an SJT had no effect on the operational validity of SJTs.⁸⁷ Overall, there is a good level of consensus among researchers that the use of SJTs represents a reliable and valid selection method across a range of occupations, including selection of medical students.^{85,88–95}

Procedural issues

Research suggests that the mode of administration may affect SJTs, with video-based SJTs having higher operational validities than equivalent paper-and-pencil SJTs.^{5,59} Similarly, different response instructions and methods of constructing alternative forms may affect the validity of the SJT selection method.^{96,97}

Acceptability

Across four studies, SJTs were rated favourably as selection tools by candidates.^{85,98–100} There is some evidence that the mode of administration may affect candidate evaluations of SJTs, with video-based SJTs rated more favourably than paper-and-pencil SJTs.⁹⁸ None of the studies identified examined the political validity or stakeholder acceptance of SJTs in medical student selection. Six studies examined the appropriateness of SJTs as a component of a wider selection process.^{85,93,101–104} The weight of evidence across these studies suggests that SJTs can usefully be incorporated into selection procedures across numerous occupational groups.

Cost-effectiveness

One study⁸⁵ concluded that there was tentative evidence for the relative cost-effectiveness of SJTs compared with other methods of assessment, although direct evidence in this area was not presented. Cost is also an important consideration in comparisons of text-based and video-based SJTs, given that video-based SJTs require significantly greater time and financial resources to develop.

Summary

There is a good level of consensus among researchers that SJTs, when properly constructed, can form a reliable, valid, cost-effective and acceptable element of medical school selection systems. SJTs are complex to develop and there is a wide range of options available in relation to item formats, instructions and scoring. When these options are calibrated appropriately, research evidence points to the strength of SJTs in medical student selection for assessing non-academic attributes.

Personality and emotional intelligence

Type of evidence

In total, 22 studies assessed the use of personality measures and six assessed the use of emotional intelligence (EI). Of the studies on personality measures, eight were longitudinal (one was a meta-analysis), five were non-systematic reviews and nine were cross-sectional (eight quantitative and one mixed methods). Two of the studies that referred to EI were longitudinal, one was a systematic review, and the other three were cross-sectional and quantitative.

Effectiveness

Despite some research finding no evidence for associations between personality traits and medical school performance,¹⁰⁵ a number of studies have found that the Big Five personality traits (openness, conscientiousness, extroversion, agreeableness and neuroticism) may correlate with various aspects of medical school performance.¹⁰⁶ Conscientiousness, for example, has also been shown to be a positive predictor of pre-clinical knowledge and examination results^{59,66,71,107} and to offer incremental validity over knowledge-based assessments.^{66,71} However, conscientiousness has also been found to be a significant negative predictor of some aspects of clinical performance,^{59,71} which demonstrates that the association between personality traits and performance in medical education and training is complex and possibly non-linear. Indeed, Ferguson et al.⁵⁹ propose that although personality research has long suggested that conscientiousness is beneficial when selecting into organisations, it has a 'dark side'; for example, facets of being methodical and dutiful may hinder the acquisition of knowledge in the clinical years of medical school. 'Dysfunctional' personality traits in medical students (including paranoid, avoidant, passive aggressive, antisocial, narcissistic and uncooperative traits) have been reported to be associated with lower academic grades.^{108,109} Considering personality assessment more broadly, it has also been demonstrated to provide incremental validity over cognitive methods in a medical school selection process.¹¹⁰ Some initial evidence exists that Western personality tests may be adapted for use in Japanese contexts,¹¹¹ although further research is required to examine the predictive validity of such adapted measures.

Two studies^{112,113} provide tentative evidence that EI may be an important characteristic in medical students that is not usually assessed by typical medical school selection methods.¹¹⁴ Other studies found no significant correlations between EI and skill in medical students,^{115,116} or other selection procedures for medical school admission.¹¹⁷ There is provisional evidence that a self-report measure of EI (the Wong and Law Emotional Intelligence Scale [WLEIS]) does not significantly correlate with measures of success in medical school, but that an ability-based measure of EI (the Mayer–Salovey–Caruso Emotional Intelligence Test [MSCEIT]) does.¹¹⁸ However, Cherry et al.¹¹⁹ concluded that there is currently insufficient evidence to support the use of EI in selection.

Procedural issues

Lievens et al.¹²⁰ suggest that the validity of personality measures in predicting medical school grades increases over the course of medical education and training. Their finding that conscientiousness is an increasing asset for medical students as their course becomes more clinical is in direct contrast to the findings reported by Ferguson et al.^{59,66} This difference may reflect the use of different populations, study designs and outcome criteria, and previous studies relying on early outcome criteria may have underestimated the predictive value of personality variables. Although there are concerns that personality tests may be 'fakeable', Hojat et al.¹⁰⁶ argue that their operational validity may be maintained by reminding respondents to reply truthfully and that intentionally false responses can be detected by a social desirability scale.

Acceptability

Evidence for the acceptability of personality assessment in medical student selection is mixed.¹²⁰ Although positive evidence on the predictive validity of personality assessment suggests that it is an appropriate and acceptable method for selecting medical students, and there is some evidence that students find personality tests acceptable for medical selection,⁴⁵ others¹²¹ have cautioned against the adoption of personality measures without considering potential future impacts on diversity in medical student personalities. No evidence was found on the acceptability of EI in medical student selection.

Cost-effectiveness

Knights and Kennedy¹⁰⁹ concluded that measures of dysfunctional personality types could be usefully and cost-effectively incorporated into medical student selection. Similarly, Powis and Rolfe¹²² gave consideration to the costs and benefits of the selection procedure at a single medical school, but did not provide any direct evidence on the cost-effectiveness of personality measures in medical student selection. No evidence was found on the cost-effectiveness of EI in medical student selection.

Summary

Taken broadly, there is a relatively high level of consensus among researchers that some domains or traits of personality are significantly positively or negatively associated with aspects of performance in medical school. However, the associations between personality domains and medical school performance are often complex, as is demonstrated by evidence that conscientiousness may be positively associated with knowledge-based assessment, but negatively associated with some clinical aspects of medical school assessment. This suggests that closer attention to the criterion constructs should also be considered when reviewing personality-based selection tools. Personality assessment can be cost-effective and may be used in combination with an interview method in which applicant responses can be probed further. Recruiters should be aware that there is a relative dearth of evidence regarding the long-term predictive validity of personality assessment beyond medical school, and that there has been some concern that personality assessment may narrow the diversity of types of individuals entering medical education and training. Research on the predictive validity of EI assessment was sparse and at a very early stage of development. The studies and reports were typically pilot studies or opinion pieces citing evidence as to why EI may represent a valuable tool in future medical student selection processes.

Interviews and multiple mini-interviews

Type of evidence

Seventy-five studies assessing the use of interviews were found. Of these, 22 were longitudinal, two were systematic reviews, four were non-systematic reviews and one was a multiple cohort study. The remaining studies were cross-sectional: four were qualitative; three used mixed methods, and 39 were quantitative.

Effectiveness

Despite some evidence to the contrary,^{14,16,33,123–130} the balance of evidence suggests that generally, the traditional interview is not a robust method of selecting medical students, and lacks predictive validity.^{4,9,28,80,131–137} Edwards et al.¹⁷ found that poorer interview performance was associated with higher medical school grade point average (GPA). The mixed findings on the effectiveness of interviews may reflect substantial differences in interview methods, which range from relatively unstructured individual interviews to highly structured panel interviews. However, Eva and Macala¹³⁸ found no difference between the reliability of interviewer ratings in unstructured and structured multiple mini-interview (MMI) stations, although behavioural indicator stations differentiated between candidates more reliably than other station types. The findings from research on MMIs tend to be more directionally consistent than those from research on traditional interviews: for example, the psychometric properties of MMIs are usually reported to be adequate.^{44,139–146} Uijtdehaage and Parker¹⁴⁶ found that the reliability of an MMI was improved by replacing an easy station with a more challenging one, and using relative, rather than absolute, ratings of candidate performance. However, Hissbach et al.¹⁴⁷ found that rater bias had a greater effect on applicant scores than systematic differences in candidate performance. There is little clarity about what is being measured within the different approaches described, although some attributes, such as communication skills, are commonly purported to be assessed by MMIs. Construct validity evidence for MMIs remains exploratory and largely inconclusive, although irrespective of design differences, the relationships between MMIs and academic measures are

small to absent.¹⁴⁵ Moreover, tightly standardised face-to-face interviews may not be comparable with scenario-based MMI stations utilising standardised role actors, and the dimensionality of MMI stations (i.e. whether MMIs can measure more than one construct per station/interview question) has been debated in the literature.¹⁴⁵ Consistent evidence of the predictive validity of MMIs is emerging from explorations of the correlation between performance on MMIs and subsequent performance on both undergraduate and postgraduate objective structured clinical examinations (OSCEs)^{143,148–152} and other examinations.^{72,153,154}

Procedural issues

Schools differ significantly in terms of the length, panel composition, structure, content and scoring methods for interviews. The differential usage of the interview method in medical student selection may underlie the mixed findings on both the reliability and validity of interviews reported above. Other research evidence suggests that candidate performance may be significantly affected by coaching.³⁰ Using interviews in a selection process also presents logistical difficulties relating to the range and type of questions¹⁵⁵ and interviewer subjectivity,^{51,143,156,157} although numerous authors report on the successful implementation of MMIs into their medical school admission processes.^{44,146} Uijtdehaage and Parker summarised that ‘implementing an MMI was feasible but a daunting task’.¹⁴⁶

Acceptability

Most research reports that applicants and interviewers tend to view the interviewing process positively,^{44,45,60,146} and there is tentative evidence that MMIs and more structured interviews are preferred over less structured methods.^{138,158} Some evidence suggests that aspiring medical students may prefer the schools that conduct interviews.¹⁵⁹ Campagna-Vaillancourt et al.¹⁴⁴ found that the majority of applicants and assessors perceived an MMI to be appropriate to assess a range of competencies and considered it to be a fair process, as well as being preferable to a traditional interview. The staged introduction of an MMI into a selection process may foster institutional acceptance of the method.¹⁶⁰ Standardised interviews can also be adapted for use in postgraduate medical selection to measure characteristics that are considered important and acceptable to both international medical graduates and interviewers.^{139,141,161}

Cost-effectiveness

The cost-effectiveness of MMIs is generally reported to be good,¹⁵⁴ although comparatively interviews are significantly more costly than machine-marked tests, and MMIs are more expensive than traditional interviews because they incur increased costs for station development and actor payments.^{145,146} Value for money may be improved by examining the number of stations in an MMI, and reducing the number of stations if reliability is not affected. However, some research suggests that increasing the number of questions or stations in MMIs increases reliability more than increasing the number of interviewers.^{143,145,162} Indeed, Roberts and colleagues estimated that to reach a Cronbach’s coefficient alpha of 0.80 for high-stakes assessment, MMIs must include 14 stations if each is manned by a single interviewer. This number could be reduced to between seven and 12 stations if each station is manned by two interviewers.¹⁴³ Alternatively, Dodson et al.¹⁶³ found that reducing the duration of MMI stations from 8 to 5 minutes conserves resources with minimal effect on applicant ranking and test reliability. Knorr and Hissbach¹⁴⁵ concluded in their systematic review that no general recommendation for the minimum number of MMI stations can be derived from the literature at present. Tiller et al.¹⁶⁴ found that cost and time savings for candidates were substantial when an MMI was conducted online via Skype rather than in person, although further research is required regarding the impact on fidelity of the lack of a face-to-face encounter.

Summary

Interviews are among the most widely used tools in selection for medical school admission. Evidence suggests that traditional interviews lack the reliability and validity that would be expected of a selection instrument in a high-stakes selection setting. Evidence also suggests that MMIs offer improved reliability and validity over traditional interview approaches. Further theory-driven research is warranted, however, in relation to the predictive and construct validity of the MMI method, particularly with respect to the constructs that can be assessed accurately (e.g. communication, critical thinking, empathy, etc.). More evidence is required regarding the appropriateness of criteria that can be assessed in interviews and should be informed by validation studies. In addition, the cost-efficiency and utility of MMIs should be evaluated, along with alternative approaches to scoring and alternative uses of scores (including any minimum threshold criteria). The use of MMIs has spread rapidly in recent years as they can be designed as a reliable selection method. However, issues surrounding the construct validity and dimensionality of MMIs remain problematic: it is critically important that schools better understand what they are seeking to measure, and actually are measuring, with this approach. The impact of the MMI on candidates (in terms of fairness, performance, coaching effects, etc.) is an outstanding practical concern that should influence design decisions such as question rotation.

Selection centres

Type of evidence

A total of seven studies assessed selection centres (SCs). One of these was longitudinal, and six were cross-sectional and quantitative.

Effectiveness

Provisional evidence indicates that SC methods may be reliable and internally valid for assessing applicants' aptitude for medicine^{165–167} and have predictive validity for performance in postgraduate training.^{168–170}

Procedural issues

Implementing an SC as part of a process for selecting medical students may be logistically complex. It requires the recruitment and training of faculty raters, and ongoing collaboration among academic and professional institutions and experts in different operational aspects of the process (including simulation, evaluation and measurement).^{167,171} Moreover, as SCs are based on a multi-trait, multimethod design, they may comprise a large number of elements in different combinations and orders, meaning that the process by which an SC is designed and administered may influence the utility of the method.

Acceptability

Provisional evidence exists that an SC for entry into postgraduate training was rated favourably by candidates and assessors.^{168–170}

Cost-effectiveness

Evidence is mixed on the cost-effectiveness of the SC method. It could be argued that SCs can offer a cost-effective method of high-volume assessment for selection into medical specialty training when balanced against the increased validity (and thus reduced extended training costs) that SCs might offer. Ziv et al.¹⁶⁷ have shown that the SC method can be expensive compared with other selection methods (approximately US\$300 per candidate) and represents a logistically complex option, although on balance they still advocate SCs for use in medical school selection. Roberts et al.¹⁷¹ investigated the feasibility of having health care staff participate in simulated scenarios as raters in order to minimise the human resources required to implement an SC. However, staff participant ratings were different from those of trained assessors and failed to achieve adequate levels of inter-

rater reliability. Nonetheless, Roberts et al.¹⁷¹ concluded that it may be viable to use other health care staff rather than trained assessors for some but not all stations.

Summary

Overall, research on the utility of SCs for medical student selection was relatively sparse. Evidence on the predictive validity of SCs for postgraduate selection is stronger, although further evidence is required to build a case for their predictive validity in medical school selection.

Table 2 summarises the findings of this review in terms of the weight of evidence for and relevance of each of the selection methods reviewed.

Table 2 An interpretation of the wider literature relating to various selection methods

Selection method	Implications of the evidence
Shortlisting methods	
Aptitude tests	Evidence is mixed on the fairness and effectiveness of aptitude tests, specifically regarding predictive validity
Academic records	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 judgement tests (SJTs)	Improved validity over other selection tools (IQ and personality tests), and can be mapped to organisational values. Although SJTs can be relatively costly to design, they can be machine-marked and can be delivered online, producing cost savings in high-volume selection
Personality assessment and emotional intelligence (EI)	Where there is a high risk for susceptibility to faking or coaching, personality assessment is best used to drive more focused questioning at interviews (rather than as a standalone 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
Final-stage selection methods	
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, trained interviewers and appropriate scoring, they can be reliable and valid. Candidates prefer interviews to other methods, although they are relatively resource-intensive.
Multiple-mini interviews (MMIs)	MMIs are relatively expensive to design and 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 tasks, presentations, interactive exercises)	SCs are relatively expensive to design and implement. Further evidence of the predictive validity of SCs in undergraduate medical selection is required

DISCUSSION

Summary of key findings

Our review of a very broad literature identifies that research into medical selection represents, to some extent, a picture of quantity over quality: a substantial number of studies are of moderate quality at best, and there are some significant gaps in the reporting and evaluation of some selection techniques. There is an over-reliance on cross-sectional study designs and a general focus on reliability estimates as indicators of quality rather than aspects of validity (a method may have high reliability but be 'reliably wrong'²⁵). Although some studies have addressed issues relating to predictive validity, very little research has explored construct validity issues (i.e. what is being

measured) and the relative cost-effectiveness of selection methods. During the 18 years covered by this review, there have been remarkably few long-term evaluation studies; however, we note that over the last 2 years there has been an increase in the amount of longitudinal evidence emerging in this area. This trend is encouraging and we anticipate that over the next 50 years increasing numbers of long-term studies will assess the predictive validity of selection methods.

There remain comparatively few studies examining selection system design overall and the relative contributions of the various selection methodologies (and the impacts of various weightings) when methods are used in combination (as is the norm in medical school selection^{172,173}). It is hard to see how substantial progress can be made without appropriately conceived long-term studies to systematically assess selection systems overall and potentially promising approaches. This is an important focus for the future research agenda in this area and this paper has sought to identify specific areas in which such work should be prioritised.

There are, however, some clear messages about the comparative reliability, validity and effectiveness of various selection methods. The academic attainment of candidates remains a common feature of most selection policies and the strength of evidence in support of it continuing to do so remains strong.

The extant evidence paints a relatively clear picture illustrating that structured interviews or MMIs, SJTs and SCs are more effective methods and generally fairer than traditional interviews, references and personal statements. Evidence is currently mixed regarding the effectiveness and fairness of aptitude tests, depending on the tool in question. This stems largely from the fact that there is no currently agreed framework that specifies what is meant by aptitude; at present tests range from assessments of 'pure' cognitive ability (e.g. the UKCAT) to academic tests (e.g. the BMAT). As such, it is difficult to systematically assess the relative contributions of different aptitude tests, and of aptitude tests within a wider selection system. The research to date largely represents isolated studies of various aptitude tests and there are no systematic reviews comparing multiple aptitude tests in a single study or cohort perhaps because the practical challenges associated with such research are substantial (i.e. in practice, it is highly unlikely that a single medical school would use multiple aptitude tests in one cohort). Similarly, more long-term validity evidence derived from the exploration of the use of personality assessments, especially in combination with other selection methods such as MMIs, is required.

The picture regarding the acceptability of various selection methods is also mixed, and may be influenced by a variety of political issues including differing stakeholder views, variations in the philosophies of both medical students and medical schools, and the ways in which the tool is implemented as part of a selection system. This area would benefit from further exploration of the reasons driving the acceptability of different selection methods.

When judging the papers in this review, it was clear that some terms cover a broad spectrum of methods: MMIs, SJTs, aptitude tests, personality assessments and SCs are measurement methods that comprise a multitude of different design parameters. For example, there are many different types of interview, even for structured interviews. Considerations of MMIs, personality tests and SJTs should note that the construction and content of the interview or test can vary significantly. Depending on the design, this may significantly alter the quality of the instrument to the extent that each needs to be individually evaluated before conclusions about its effectiveness can be reached. Although results from meta-analytic studies can indicate the quality of different selection methods in general, local validation studies are required to determine the effectiveness of any given selection system.

Implications for theory

A persistent problem with selection research relates to the issue of which outcomes we are trying to predict by using various selection methods.⁵⁹ For example, to illustrate this criterion problem, when exploring the association between conscientiousness and performance outcomes, we find mixed results when examining outcomes relating to early examination performance in medical school and performance within clinical practice in later years. Furthermore, our review also highlights that outcome measures used to evaluate selection methods most often focus on indicators of attainment and maximal performance (e.g. medical school achievements, performance in licensure examinations) rather than indicators relating to clinical practice and typical (day-to-day) in-role job performance.

The (few) longitudinal predictive validity studies available often lack sufficient detail regarding the target outcome variables with which to interpret results. In judging the evidence for the relative accuracy of selection methods, it becomes apparent that a clear framework of outcome criteria with which to interpret the research evidence and compare selection methods, both individually, and within a selection system, has yet to be established; future research should urgently address this gap in our understanding. In addition, evidence regarding the effectiveness of some methods has focused predominantly on the predictive validity of the tool, rather than on assessing precisely what different methods are measuring (i.e. construct validity); this raises the question of how a method can be considered to add value to a selection system if the constructs it is measuring are unknown. This is particularly the case for MMI research, in which, despite the method's increasing popularity in recent years, there is a lack of consistency regarding the attributes selectors are using MMIs to assess for and, relatedly, evidence regarding construct validity remains inconclusive.

It is clear that indicators of competence for entrance to medical training and practice are likely to be different at different points in a medical career; thus, applicants are judged on multiple selection criteria depending on the specific role, which may include varying combinations of academic and non-academic indicators of aptitude. A factor may be identified as an important predictor for undergraduate training, but may actually hinder some aspects of performance in clinical practice.^{59,66} As such, different selection methods may predict differently at different stages: for example, an SJT may be less predictive of performance in the early years at medical school (which tends to be more academically-focused), but significantly more predictive of performance outcomes when trainees enter clinical practice.^{28,174} A major challenge within medicine is to integrate the research evidence to inform the design of selection systems that are reliable and valid (and weighted appropriately) from undergraduate selection through to selection for specialty training after many years of education, for both academic and non-academic qualities. This requires a clearer, theoretically relevant taxonomy of desirable outcomes which might range from academically oriented variables such as examination performance, through to variables relating to clinical practice and job performance indicators as judged by supervisors, peers and, ideally, patients (i.e. multi-source feedback). Hence, there is a need for more theoretically driven, future-oriented research aimed at identifying what a 'competent' physician is at the various stages of training and practice. This will allow researchers and practitioners to move towards crafting a unified taxonomy of performance indicators which may be used as markers in short- and long-term predictive validity studies of selection methods. For example, some researchers suggest that from undergraduate selection onwards, medical students should be selected in on the basis of academic attainment and selected out on the basis of non-academic skills and attributes.¹⁷⁵ It could be argued that non-academic attributes and skills should therefore play a much larger role in postgraduate selection and the weighting of these may differ depending on the specialty. For example, research from job analysis studies shows that empathy and communication are weighted more heavily for selection into general practice¹⁷⁶ and paediatrics, whereas vigilance and situational awareness carry more weight in anaesthesia.¹⁷⁷

Implications for practice

A challenge in practice to date has concerned the measurement of important non-academic attributes (e.g. empathy, integrity) reliably at the point of selection, using an appropriate combination of methods that balance efficiency, cost-effectiveness, procedural issues and stakeholder acceptability. Our review shows that SJTs and MMIs are more valid predictors of inter- and intrapersonal (non-academic) attributes than personal statements or references.

Situational judgement tests (SJTs) and MMIs may be complementary: whereas SJTs can measure a broader range of constructs efficiently as they can be machine-marked, MMIs, by contrast, involve a face-to-face encounter. Although expensive, structured interviews (including MMIs) allow applicant responses to be probed further and in more depth. Here, results from personality assessments could add significant value if they were to be used alongside a structured interview. Subjects for future research and practice should be the design and long-term evaluation of effective and scalable methods to assess non-academic attributes accurately, and an exploration of optimal combinations of tools. The advent of the use of technology in interviews may have the potential to revolutionise how interviews are conducted; completing interviews online may represent significant cost and time savings for applicants who are geographically dispersed.

At present, the picture for aptitude tests and cognitive factors is less clear as a result of the large number of aptitude tests and the differences between those that are currently available, the diverse outcome measures against which performance on aptitude tests is compared (to assess validity, see the 'criterion problem' discussed above), the multiple ways in which aptitude tests are implemented, and the mixed nature of the evidence on the effectiveness of aptitude testing. There is also some evidence that some aptitude tests may favour certain types of candidate,⁴⁶ which may have unfavourable implications for fairness and widening access to medicine.

However, more high-quality research is required to explore such findings in further detail, and to indicate how individual aptitude tests should be used most prudently and fairly with reference to specific outcome criteria.

Interpreting the breadth of the currently available literature is challenging: although some practitioners feel there is insufficient evidence to evaluate selection methods, others argue that there is so much evidence available that it is overwhelming to try to collate it to identify which selection methods are 'best'. The challenges of interpreting and applying evidence of the relative acceptability, cost-effectiveness, practical issues and effectiveness (including reliability and validity) of selection methods include the relative lack of longitudinal data, lack of an agreed-upon framework of outcome criteria, and institutional differences (including in available resources, curricula and philosophies of what a high-performing medical student is considered to be). Kreiter and Axelson² acknowledge that the complexity of admissions goals may also be an obstacle to evidence-based progress in medical school admissions because concerns regarding social justice, educational equality, health care and political outcomes are broad and frequently competing. When judging the quality and effectiveness of selection methods, it is noteworthy that some criteria may compete with one another. For example, the stakeholder acceptability of referees' reports in selection is generally high, but the evidence for their validity is poor. Similarly, regarding other criteria, the evidence for the validity of SCs is high, but they are relatively costly to implement. In this respect, when judging the quality and effectiveness of different selection methods, medical schools and employers may choose to weight different features depending on the context within which the selection system is operating. This review was intended to synthesise the literature so that the reader may gain an understanding of current research regarding the strengths and limitations of each method, rather than to prescribe a single, best approach. Ultimately, the aim is to design efficient, acceptable and fair methods that can be scaled up for high-volume use. This review highlights that at present there is not sufficient evidence to suggest that any one selection method currently in use meets all of these criteria.

We propose that a key implication for practice derived from the considerations above, and for the suggested research agenda outlined below, refers to the necessity for collaborative international studies involving multiple sites to gather and analyse high-quality, longitudinal data about the effectiveness, cost-efficiency, issues of implementation and stakeholder acceptability of selection methods. In so doing there is an opportunity to gain practical, in-depth and long-term knowledge about the relative efficiency of different selection methods. A common central concern for any selection tool is susceptibility to coaching. Research over the last 10 years has increasingly focused on this issue, probably because there has been increasing emphasis on how to validly assess non-academic attributes in selection for medical education. In particular, personal statements are at significant risk of being influenced by coaching, or indeed of being written by somebody other than the applicant; a brief online search reveals a large number of companies internationally that sell pre-written personal statements. With regard to SJTs, recent studies have found no effects of commercial coaching on SJT scores or the predictive validity of SJTs.^{87,178} However, ongoing research is required to assess the coachability of the full range of non-academic selection tools in greater depth.

In future, researchers may also want to explore how assessment data can be used to direct educational interventions and create more tailored education programmes to accelerate time to competence. Such assessment data may also be used for diagnostic purposes and complemented by high-fidelity interventions. Interventions could provide training in the professionalism skills necessary to work in multidisciplinary health care teams.

Scoping a future research agenda It is clear from our review that it is challenging to draw firm conclusions regarding the relative strength of the different tools given the variety in the quality and design of the currently available research evidence: at present there are insufficient data, and medical education providers' agendas are too diverse, to propose a fully comprehensive framework for international best practice in medical selection methods. Hence, here we outline a possible future research agenda, which may help to strengthen the evidence about each selection tool, in order to progress researchers' and practitioners' knowledge towards a framework for best practice in medical school selection methods.

Although the literature on selection methods is large, there exist many uncharted territories for further research. There is a clear need for well-planned studies focusing on the long-term follow-up of medical students, tracking students from admission through to assessments in more senior training posts in clinical practice, at the point of licensure and beyond. This review clearly highlights the lack of evidence available for medical schools and employers to use in making decisions about which selection tools to use, in which combinations, and what individual weightings to apportion to each tool used.

Within the broader sphere of issues of fairness in selection, more research exploring issues of widening access and diversity is required, whether it refers to race, ethnicity or social class, as this remains a challenge within medical school admissions globally, and it is becoming increasingly important politically to reflect society within the health care professions. ^{179,180} Indicators of socio-demography pertinent to each country often reflect the same underpinning socio-economic bias, which presents either a barrier to entry to study or reduced chances of successful application. The preceding literature review highlights a paucity of educational research of sufficient quality and type to adequately assess the impact of a variety of selection tools upon widening access robustly. For example, O'Neill et al.¹⁸¹ found no significant effect of selection method on social diversity in the medical student population, and suggest that the attraction of a sufficiently diverse applicant pool is more important for widening access than which selection tool is used. Therefore, only tentative conclusions can be drawn. It is likely that some selection tools are more sensitive to social bias than

others, but more definitive data are required. For example, initial evaluation of SJTs at entry to medical school level confirms that applicants' performance at testing does not follow the usual socio-economic trends, as do tests of academic attainment;¹⁸² further research is required to explain why this might be the case. There is also initial evidence to suggest that MMIs may be equitable with regard to the demographic status of applicants.¹⁴⁴

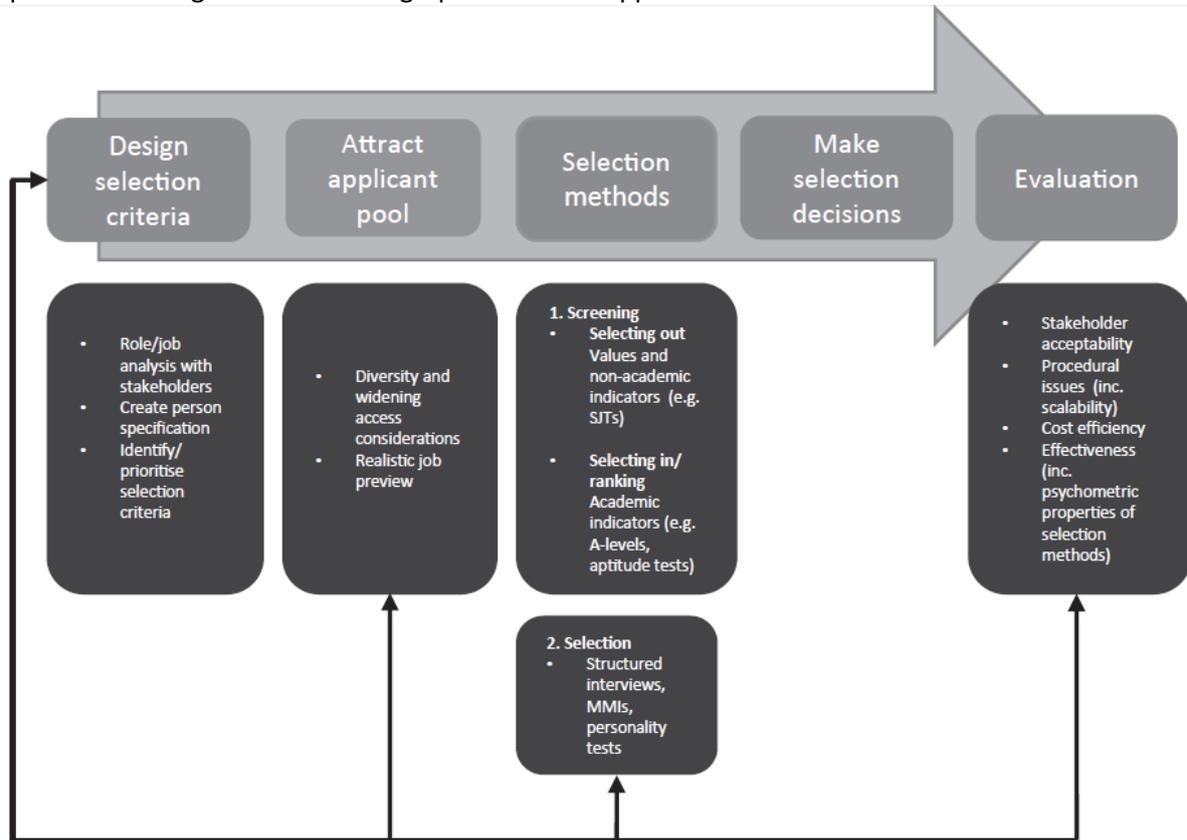


Figure 2 Design and evaluation of selection systems. MMIs = multiple mini-interviews; SJTs = situational judgement tests

Reports which address aptitude tests, such as the UKCAT, have shown that institutions with selection policies that favour the use of such a test tend to make more offers to applicants from disadvantaged backgrounds, and that the aptitude test itself is less sensitive than traditional measures of academic attainment to some socio-economic markers, such as school type.^{50,183} Whereas traditional markers of prior educational attainment have been called the 'academic backbone' of medical education because they are highly predictive of subsequent performance both at medical school and beyond, there is a need to explore how 'contextual data' can be used to allow the social and educational backgrounds of applicants to be taken into consideration alongside their educational achievements. Prior academic attainment is clearly still an important component of the medical selection process, but care must be taken to ensure that it is included in a way that guarantees it does not represent a barrier to candidates from disadvantaged groups.

A key criticism of selection research is that there is a distinct lack of theory-driven studies that examine issues related to validity and the constructs being measured and that, more broadly, acknowledge contemporary models of adult intellectual development and skill acquisition, or attempt to integrate cognitive and non-cognitive factors.^{172,173} The term 'noncognitive' is in itself problematic as it arguably implies 'not thinking'; future research must also look towards more theoretical underpinnings, drawing on not just psychometric approaches but also theoretical models of adult intellectual functioning, personality, values, and individual differences. For example, there

has been little previous research exploring how to assess values as part of recruitment to the health care professions, yet compassion and benevolence are important for any health care professional to facilitate the provision of high quality care and patient outcomes, and hence new research literature in this area is now emerging.¹⁸⁴ Only by exploring the theoretical underpinnings will research in selection progress to enable a richer understanding of how personality, aptitude, interest, values and motivation interact to define areas of competence and career choice.

In summary, we propose the following priorities for a future research agenda over the next 50 years in order to enable schools and employers to make evidence-based decisions about which selection tools to use and why:

- 1** longitudinal research exploring predictive validity and following students throughout the course of their careers within education, training and practice;
- 2** research enabling greater understanding of how selection tools may impact on widening access and diversity agendas, and
- 3** theory-driven studies of the construct validity of both academically and non-academically oriented selection methods and selection systems that will help us to understand what we are assessing for in both the short and long terms.

Crucially, future research must consider selection tools within the context of wider selection systems, rather than viewing each tool in isolation, because some tools are predictive of performance in clinical practice but not during the first few years of undergraduate medical degrees, and some tools may be complementary for ‘selecting in’ at the top end of performers and ‘selecting out’ at the bottom end. Figure 2 provides a model of the design and evaluation of selection systems that outlines the interrelated nature of the different stages of a selection system. As we have discussed, further research on the various weightings to be applied to each stage or selection method is required, as well as feedback from every stage at each iteration of the selection process. Finally, we propose that the following five considerations will be integral in shaping the direction of medical education research over the next 50 years:

- 1.** Medical school admissions will remain highly competitive. The prestige of being a physician is likely to continue to drive a high applicant-to-selection ratio in medical school selection internationally over the next 50 years. However, this is unlikely to be true in all postgraduate specialties; some medical career pathways may be perceived to be of higher status and will therefore be more competitive than others. Medical selection may become part of a process to facilitate recruitment into areas of most need. This may, in turn, require varying emphasis on selection for specific attributes and competencies: one size is unlikely to fit all.
- 2.** There will be an increased focus on, and value of, non-academic attributes and skills in medical selection, aligned with what wider society wishes from its physicians. The role of the physician’s own wellbeing and resilience, and how these can best be selected for, then supported and developed, will be of increasing importance. Trainees’ expectations of their work–life balance will also be integral to medical selection over the next 50 years. Consideration must be given during selection to the discourse around how we encourage new generations of medical students to expend discretionary effort in future. This is strongly related to:
- 3.** A growing focus on capability to lead multidisciplinary teams, and building a culture of ‘everyday’ innovation in an environment of reduced resources.
- 4.** Rather than a focus on just one or two people in a team, who are touted as the ‘innovators’, there is likely to be an increased onus on all health care professionals to innovate and provide leadership in order to engage multi-professional teams and to continue to deliver high-quality and

compassionate care in a climate of ongoing health care spending cuts.^{185,186} This may represent a significant change in how applicants to medical education are selected.

This, in turn, relates to:

5. a focus on attracting a wider selection pool and recruiting a more diverse workforce, reflecting a philosophical shift towards acknowledging that non-traditional students may be able to align themselves with patients from diverse backgrounds and also contribute to the education of their peers by acting to challenge the current medical culture.^{187,188} Bringing such 'nontraditional' applicants into the health care system may promote, and indeed necessitate, innovative working practices. However, as we have discussed elsewhere,¹⁸⁰ there is currently a multitude of unanswered questions on how this may be best implemented and how outcomes can be measured in a reliable and valid way.

Strengths and limitations

A key strength of our review is that we have collated and synthesised the breadth of research evidence collected over the last 18 years in order to draw conclusions on five key evaluation criteria for medical education selection methods. We also identify current gaps in understanding and theory, and outline a future research agenda that aims to address these areas. Attempting to summarise our conclusions from the large number of studies reviewed in Table S3 (on-line) naturally runs the risk of simplifying some of the intricacies of the studies and the nuances of their findings. We therefore encourage our readers to consider the original source should they wish to gain a fuller picture of each study's context, rationale, methodology and findings. Nonetheless, we feel that Table S3 provides a valuable resource through which the reader can identify key papers and navigate the sizeable and diverse literature base.

Contributors: all authors made substantial contributions to the conception, analysis and design of the work. AK, FC and FP led the data acquisition and analysis. FP prepared the first draft of this paper and all authors contributed to the critical revision of the document through several iterations. All authors approved the final manuscript for publication.

Acknowledgements: the authors thank the UK General Medical Council (GMC) for commissioning an initial rapid review of the literature on selection and widening access to medicine in 2013. Further thanks are due to the UK Medical Schools Council (MSC) for commissioning a significantly updated review of selection methods in 2014, funding for which was provided by Health Education England (HEE) and the Office for Fair Access (OFFA). Working together on both of these projects encouraged us to further develop our ideas and produce an updated systematic review for publication in 2015. We also thank those who contributed to the original project funded by the GMC, notably John McLachlan, Member of the Centre for Medical Education Research, Durham University, and Emma Dunlop, Medical Admissions, University of Aberdeen.

Funding: funding was provided by HEE and the OFFA.

Conflicts of interest: FP, FC and AK provide advice on selection methodology to HEE and UKCAT through Work Psychology Group Ltd. SN is a board member and research panel lead for UKCAT. JC is a member of the UKCAT research group.

Ethical approval: this study was approved by the University of Aberdeen's College of Life Sciences and Medicine Ethics Review Board.

REFERENCES

- 1 Patterson F, Lievens F, Kerrin M, Zibarras L, Carette B. Designing selection systems for medicine: the importance of balancing predictive and political validity in high-stakes selection contexts. *Int J Select Assess* 2012;20:486–96.
- 2 Kreiter CD, Axelson RD. A perspective on medical school admission research and practice over the last 25 years. *Teach Learn Med* 2013;25 (Suppl):50–6.
- 3 Coates H. Establishing the criterion validity of the Graduate Medical School Admissions Test (GAMSAT). *Med Educ* 2008;42 (10):999–1006.
- 4 Trost G, Nauels HU, Klieme E. The relationship between different criteria for admission to medical school and student success. *Assess Educ Princ Pol Pract* 1998;5:247–54.
- 5 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? *Ir Med J* 2010;103:300–2.
- 6 Patterson F, Ferguson E. *Selection for Medical Education and Training*. Oxford: Wiley Online Library 2010.
- 7 Ferguson E, James D, Madeley L. Factors associated with success in medical school: systematic review of the literature. *BMJ* 2002;324:952–7.
- 8 Milburn A. *University Challenge: How Higher Education can Advance Social Mobility*. A progress report by the Independent Reviewer on Social Mobility and Child Poverty. London: Cabinet Office 2012.
- 9 Prideaux D, Roberts C, Eva K, Centeno A, McCrorie P, McManus C, Patterson F, Powis D, Tekian A, Wilkinson D. Assessment for selection for the health care professions and specialty training: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach* 2011;33:215–23.
- 10 Petticrew M, Roberts H. *Systematic reviews in the social sciences: a practical Guide*. Oxford: John Wiley & Sons 2008.
- 11 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. *Acad Med* 2010;85:980–7.
- 12 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. *Acad Med* 2013;88:666–71.
- 13 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. *Teach Learn Med* 2002;14:98–103.
- 14 Peskun C, Detsky A, Shandling M. Effectiveness of medical school admissions criteria in predicting residency ranking four years later. *Med Educ* 2007;41 (1):57–64.
- 15 Mercer A, Crotty B, Alldridge L, Le L, Vele V. GAMSAT: a 10-year retrospective overview, with detailed analysis of candidates' performance in 2014. *BMC Med Educ* 2015;15:31.

- 16 Puddey IB, Mercer A. Predicting academic outcomes in an Australian graduate entry medical programme. *BMC Med Educ* 2014;14:31.
- 17 Edwards D, Friedman T, Pearce J. Same admissions tools, different outcomes: a critical perspective on predictive validity in three undergraduate medical schools. *BMC Med Educ* 2013;13:173.
- 18 Poole P, Shulruf B, Rudland J, Wilkinson T. Comparison of UMAT scores and GPA in prediction of performance in medical school: a national study. *Med Educ* 2012;46 (2):163–71.
- 19 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 Med Educ* 2014;14:88.
- 20 McManus IC, Dewberry C, Nicholson S, Dowell JS. The UKCAT-12 study: educational attainment, aptitude test performance, demographic and socioeconomic contextual factors as predictors of first year outcome in a cross-sectional collaborative study of 12 UK medical schools. *BMC Med* 2013;11:244.
- 21 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 Med Educ* 2014;14:116.
- 22 Wright SR, Bradley PM. Has the UK Clinical Aptitude Test improved medical student selection? *Med Educ* 2010;44 (11):1069–76.
- 23 Bell JF. The case against the BMAT: not withering but withered? *BMJ* 2005;331:555.
- 24 Emery JL, Bell JF, Vidal Rodeiro CL. The BioMedical Admissions Test for medical student selection: issues of fairness and bias. *Med Teach* 2011;33:62–71.
- 25 Albishri JA, Aly SM, Alnemary Y. Admission criteria to Saudi medical schools. Which is the best predictor for successful achievement? *Saudi Med J* 2012;33:1222–6.
- 26 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. *Acad Med* 2007;82:100–6.
- 27 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 Med Educ* 2010;10:1–9.
- 28 Wilkinson D, Zhang J, Byrne GJ, Luke H, Ozolins IZ, Parker MH, Peterson RF. Medical school selection criteria and the prediction of academic performance. *Med J Aust* 2008;188:349–54.
- 29 Griffin B, Yeomans ND, Wilson IG. Students coached for an admission test perform less well throughout a medical course. *Intern Med J* 2013;43:927–32.
- 30 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 Med Educ* 2013;13: 159.

- 31 Poole P, Shulruf B. Shaping the future medical workforce: take care with selection tools. *J Prim Health Care* 2013;5:269–75.
- 32 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 Med Educ* 2014;14:48.
- 33 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 Med Educ* 2014;14:86.
- 34 Wilkinson D, Zhang J, Parker M. Predictive validity of the Undergraduate Medicine and Health Sciences Admission Test for medical students’ academic performance. *Med J Aust* 2011;194:341–4.
- 35 McManus IC, Ferguson E, Wakeford R, Powis D, James D. Predictive validity of the Biomedical Admissions Test: an evaluation and case study. *Med Teach* 2011;33:53–7.
- 36 McManus IC, Ferguson E, Wakeford R, Powis D, James D. Response to comments by Emery and Bell, *Medical Teacher* 33 (1): (this issue). *Med Teach* 2011;33:60–1.
- 37 Al-Rukban MO, Munshi FM, Abdulghani HM, Al-Hoqail I. The ability of the pre-admission criteria to predict performance in a Saudi medical school. *Saudi Med J* 2010;31:560–4.
- 38 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. *Med Teach* 2013;35:e1123–9.
- 39 Adam J, Dowell J, Greatrix R. Use of UKCAT scores in student selection by UK medical schools, 2006– 2010. *BMC Med Educ* 2011;11:98.
- 40 Albanese MA, Farrell P, Dottl S. Statistical criteria for setting thresholds in medical school admissions. *Adv Health Sci Educ Theory Pract* 2005;10:89–103.
- 41 Hissbach J, Klusmann D, Hampe W. Dimensionality and predictive validity of the HAM-Nat, a test of natural sciences for medical school admission. *BMC Med Educ* 2011;11:1–11.
- 42 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. *Acad Med* 2010;85 (Suppl):64–7.
- 43 Cleland JA, French FH, Johnston PW. A mixed-methods study identifying and exploring medical students’ views of the UKCAT. *Med Teach* 2011;33:244–9.
- 44 Kelly M, Dowell J, Husbands A, Kropmans T, Jackson AE, Dunne F, O’Flynn S, Newell J, Murphy AW. Can multiple mini interviews work in an Irish setting? A feasibility study *Ir Med J* 2014;107:201–2.
- 45 Stevens L, Kelly ME, Hennessy M, Last J, Dunne F, O’Flynn S. Medical students views on selection tools for medical school – a mixed methods study. *Ir Med J* 2014;107:229–31.
- 46 Aldous CJ, Leeder SR, Price J, Sefton AE, Teubner JK. A selection test for Australian graduate-entry

medical schools. *Med J Aust* 1997;166:247–50.

47 Griffin B, Harding DW, Wilson IG, Yeomans ND. Does practice make perfect? The effect of coaching and retesting on selection tests used for admission to an Australian medical school. *Med J Aust* 2008;189:270–3.

48 Lambe P, Waters C, Bristow D. The UK Clinical Aptitude Test: is it a fair test for selecting medical students? *Med Teach* 2012;34:e557–65.

49 Winegarden B, Glaser D, Schwartz A, Kelly C. MCAT Verbal Reasoning score: less predictive of medical school performance for English language learners. *Med Educ* 2012;46 (9):878–86.

50 Tiffin PA, McLachlan JC, Webster L, Nicholson S. Comparison of the sensitivity of the UKCAT and A levels to sociodemographic characteristics: a national study. *BMC Med Educ* 2014;14:7.

51 Griffin BN, Wilson IG. Interviewer bias in medical student selection. *Med J Aust* 2010;193:343–6.

52 O'Flynn S, Fitzgerald T, Mills A. Modelling the impact of old and new mechanisms of entry and selection to medical school in Ireland: who gets in? *Ir J Med Sci* 2013;182:421–7.

53 Bhatti MA, Anwar M. Does entry test make any difference on the future performance of medical students? *J Pak Med Assoc* 2012;62:664–8.

54 Cohen-Schotanus J, Muijtjens AMM, Reinders JJ, Agsteribbe J, van Rossum HJM, van der Vleuten CPM. The predictive validity of grade point average scores in a partial lottery medical school admission system. *Med Educ* 2006;40 (10):1012–19.

55 Kreiter CD, Kreiter Y. A validity generalisation perspective on the ability of undergraduate GPA and the medical college admission test to predict important outcomes. *Teach Learn Med* 2007;19:95–100.

56 Lumb AB, Vail A. Comparison of academic, application form and social factors in predicting early performance on the medical course. *Med Educ* 2004;38 (9):1002–5.

57 Luqman M. Relationship of academic success of medical students with motivation and pre-admission grades. *J Coll Physicians Surg Pak* 2013;23:31–6.

58 McManus IC, Smithers E, Partridge P, Keeling A, Fleming PR. A levels and intelligence as predictors of medical careers in UK doctors: 20-year prospective study. *BMJ* 2003;327:139–42.

59 Ferguson E, Semper H, Yates J, Fitzgerald JE, Skatova A, James D. The 'dark side' and 'bright side' of personality: when too much conscientiousness and too little anxiety are detrimental with respect to the acquisition of medical knowledge and skill. *PLoS One* 2014;9:e88606.

60 Kelly ME, Dowell J, Husbands A, Newell J, O'Flynn S, Kropmans T, Dunne FP, Murphy AW. The fairness, predictive validity and acceptability of multiple mini interview in an internationally diverse student population – a mixed methods study. *BMC Med Educ* 2014;14:267.

61 Mufti T, Qayum I. Rehman Medical College admission criteria as an indicator of students' performance in university professional examinations. *J Ayub Med Coll Abbottabad* 2013;26:564–7.

- 62 Schripsema NR, van Trigt AM, Borleffs JCC, Cohen-Schotanus J. Selection and study performance: comparing three admission processes within one medical school. *Med Educ* 2014;48 (12):1201–10.
- 63 McManus IC, Dewberry C, Nicholson S, Dowell JS, Woolf K, Potts HW. Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student selection: meta-regression of six UK longitudinal studies. *BMC Med* 2013;11:243.
- 64 O'Neill L, Hartvigsen J, Wallstedt B, Korsholm L, Eika B. Medical school dropout – testing at admission versus selection by highest grades as predictors. *Med Educ* 2011;45 (11):1111–20.
- 65 Urlings-Strop LC, Stegers-Jager KM, Stijnen T, Themmen AP. Academic and non-academic selection criteria in predicting medical school performance. *Med Teach* 2013;35:497–502.
- 66 Ferguson E, James D, O'Hehir F, Sanders A, McManus IC. Pilot study of the roles of personality, references, and personal statements in relation to performance over the five years of a medical degree. *BMJ* 2003;326:429–32.
- 67 Tektas OY, Fiessler C, Mayr A, Neuhuber W, Paulsen F. Correlation of high school exam grades with study success at a German medical school. *J Contemp Med Educ* 2013;1:157–62.
- 68 McManus IC, Powis DA, Wakeford R, Ferguson E, James D, Richards P. Intellectual aptitude tests and A levels for selecting UK school leaver entrants for medical school. *BMJ* 2005;331:555–9.
- 69 McManus IC, Woolf K, Dacre J. Even one star at A level could be 'too little, too late' for medical student selection. *BMC Med Educ* 2008; 8:16.
- 70 Oosterveld P, ten Cate O. Generalisability of a study sample assessment procedure for entrance selection for medical school. *Med Teach* 2004;26:635–9.
- 71 Ferguson E, Sanders A, O'Hehir F, James D. Predictive validity of personal statements and the role of the five-factor model of personality in relation to medical training. *J Occup Organ Psychol* 2000;73:321–44.
- 72 Husbands A, Dowell J. Predictive validity of the Dundee multiple mini-interview. *Med Educ* 2013;47 (7):717–25.
- 73 Wouters A, Bakker AH, van Wijk IJ, Croiset G, Kusurkar RA. A qualitative analysis of statements on motivation of applicants for medical school. *BMC Med Educ* 2014;14:200.
- 74 White J, Brownell K, Lemay JF, Lockyer JM. 'What do they want me to say?' The hidden curriculum at work in the medical school selection process: a qualitative study. *BMC Med Educ* 2012;12:17.
- 75 White JS, Lemay JF, Brownell K, Lockyer J. 'A chance to show yourself' – how do applicants approach medical school admission essays? *Med Teach* 2011;33:e541–8.
- 76 Elam CL, Johnson MM. The effect of a rolling admission policy on a medical school's selection of applicants. *Acad Med* 1997;72:644–6.

- 77 Hanson MD, Dore KL, Reiter HI, Eva KW. Medical school admissions: revisiting the veracity and independence of completion of an autobiographical screening tool. *Acad Med* 2007;82 (Suppl):8–11.
- 78 Parry J, Mathers J, Stevens A, Parsons A, Lilford R, Spurgeon P, Thomas H. Admissions processes for five year medical courses at English schools: review. *BMJ* 2006;332:1005–9.
- 79 Kumwenda B, Dowell J, Husbands A. Is embellishing UCAS personal statements accepted practice in applications to medicine and dentistry? *Med Teach* 2013;35:599–603.
- 80 Benbassat J, Bauml R. Uncertainties in the selection of applicants for medical school. *Adv Health Sci Educ Theory Pract* 2007;12:509–21.
- 81 Poole PJ, Moriarty HJ, Wearn AM, Wilkinson TJ, Weller JM. Medical student selection in New Zealand: looking to the future. *N Z Med J* 2009;122:88–100.
- 82 DeZee KJ, Magee CD, Rickards G, Artino AR Jr, Gilliland WR, Dong T, McBee E, Paolino N, Cruess DF, Durning SJ. What aspects of letters of recommendation predict performance in medical school? Findings from one institution *Acad Med* 2014;89:1408–15.
- 83 Stedman JM, Hatch JP, Schoenfeld LS. Letters of recommendation for the predoctoral internship in medical schools and other settings: do they enhance decision making in the selection process? *J Clin Psychol Med Settings* 2009;16:339–45.
- 84 Cullen MJ, Sackett PR, Lievens F. Threats to the operational use of situational judgement tests in the college admission process. *Int J Select Assess* 2006;14:142–55.
- 85 Lievens F, Peeters H, Schollaert E. Situational judgement tests: a review of recent research. *Pers Rev* 2008;37:426–41.
- 86 Rostom H, Watson R, Leaver L. Situational judgement tests: the role of coaching. *Med Educ* 2013;47 (2):219.
- 87 Simon E, Walsh K, Paterson-Brown F, Cahill D. Does a high ranking mean success in the Situational Judgement Test? *Clin Teach* 2015;12 (1):42–5.
- 88 Cabrera MAM, Nguyen NT. Situational judgement tests: a review of practice and constructs assessed. *Int J Select Assess* 2001;9:103–13.
- 89 Christian MS, Edwards BD, Bradley JC. Situational judgement tests: constructs assessed and a meta-analysis of their criterion-related validities. *Pers Psychol* 2010;63:83–117.
- 90 Hansel M, Klupp S, Graupner A, Dieter P, Koch T. Dresden Faculty selection procedure for medical students: what impact does it have, what is the outcome? *GMS Z Med Ausbild* 2010;27:Doc25.
- 91 Lievens F. Adjusting medical school admission: assessing interpersonal skills using situational judgement tests. *Med Educ* 2013;47 (2):182–9.
- 92 Lievens F, Buyse T, Sackett PR. The operational validity of a video-based situational judgement test for medical college admissions: illustrating the importance of matching predictor and criterion construct domains. *J Appl Psychol* 2005;90:442–52.

- 93 Patterson F, Carr V, Zibarras L, Burr B, Berkin L, Plint S, Irish B, Gregory S. New machine-marked tests for selection into core medical training: evidence from two validation studies. *Clin Med* 2009;9:417–20.
- 94 Libbrecht N, Lievens F, Carette B, Cote S. Emotional intelligence predicts success in medical school. *Emotion* 2014;14:64–73.
- 95 Patterson F, Ashworth V, Kerrin M, O’Neill P. Situational judgement tests represent a measurement method and can be designed to minimise coaching effects. *Med Educ* 2013;47(2):220–1.
- 96 Lievens F, Sackett PR. Situational judgement tests in high-stakes settings: issues and strategies with generating alternate forms. *J Appl Psychol* 2007;92:1043–55.
- 97 McDaniel MA, Hartman NS, Whetzel DL, Grubb WL. Situational judgement tests, response instructions, and validity: a meta-analysis. *Pers Psychol* 2007;60:63–91.
- 98 Chan D, Schmitt N. Situational judgement and job performance. *Hum Perform* 2002;15:233–54.
- 99 Koczwara A, Patterson F, Zibarras L, Kerrin M, Irish B, Wilkinson M. Evaluating cognitive ability, knowledge tests and situational judgement tests for postgraduate selection. *Med Educ* 2012;46(4):399–408.
- 100 Plint S, Patterson F. Identifying critical success factors for designing selection processes into postgraduate specialty training: the case of UK general practice. *Postgrad Med J* 2010;86:323–7.
- 101 Ahmed H, Rhydderch M, Matthews P. Can knowledge tests and situational judgement tests predict selection centre performance? *Med Educ* 2012;46(8):777–84.
- 102 Clevenger J, Pereira GM, Wiechmann D, Schmitt N, Harvey VS. Incremental validation of situational judgement tests. *J Appl Psychol* 2001;86:410–7.
- 103 O’Connell MS, Hartman NS, McDaniel MA, Grubb WL, Lawrence A. Incremental validity of situational judgement tests for task and contextual job performance. *Int J Select Assess* 2007;15:19–29.
- 104 Patterson F, Baron H, Carr V, Plint S, Lane P. Evaluation of three short-listing methodologies for selection into postgraduate training in general practice. *Med Educ* 2009;43(1):50–7.
- 105 Haight SJ, Chibnall JT, Schindler DL, Slavin SJ. Associations of medical student personality and health/wellness characteristics with their medical school performance across the curriculum. *Acad Med* 2012;87:476–85.
- 106 Hojat M, Erdmann JB, Gonnella JS. Personality assessments and outcomes in medical education and the practice of medicine: AMEE Guide No. 79. *Med Teach* 2013;35:e1267–301.
- 107 Lievens F, Coetsier P, De Fruyt F, De Maeseneer J. Medical students’ personality characteristics and academic performance: a five-factor model perspective. *Med Educ* 2002;36(11):1050–6.

- 108 Knights JA, Kennedy BJ. Medical school selection: screening for dysfunctional tendencies. *Med Educ* 2006;40 (11):1058–64.
- 109 Knights JA, Kennedy BJ. Medical school selection: impact of dysfunctional tendencies on academic performance. *Med Educ* 2007;41 (4):362–8.
- 110 Chan-Ob T, Boonyanaruthee V. Medical student selection: which matriculation scores and personality factors are important? *J Med Assoc Thai* 1999;82:604–10.
- 111 Fukui Y, Noda S, Okada M, Mihara N, Kawakami Y, Bore M, Munro D, Powis D. Trial use of the Personal Qualities Assessment (PQA) in the entrance examination of a Japanese medical university: similarities to the results in western countries. *Teach Learn Med* 2014;26:357–63.
- 112 Carrothers RM, Gregory SW Jr, Gallagher TJ. Measuring emotional intelligence of medical school applicants. *Acad Med* 2000;75:456–63.
- 113 Edwards JC, Elam CL, Wagoner NE. An admission model for medical schools. *Acad Med* 2001;76:1207–12.
- 114 Bore M, Munro D, Powis D. A comprehensive model for the selection of medical students. *Med Teach* 2009;31:1066–72.
- 115 Carr SE. Emotional intelligence in medical students: does it correlate with selection measures? *Med Educ* 2009;43 (11):1069–77.
- 116 Lin DT, Kannappan A, Lau JN. The assessment of emotional intelligence among candidates interviewing for general surgery residency. *J Surg Educ* 2013;70:514–21.
- 117 Leddy JJ, Moineau G, Puddester D, Wood TJ, Humphrey-Murto S. Does an emotional intelligence test correlate with traditional measures used to determine medical school admission? *Acad Med* 2011;86 (Suppl):39–41.
- 118 Brannick M, Grichanik M, Nazian S, Wahi M, Goldin S. Emotional intelligence and medical school performance: a prospective multivariate study. *Med Sci Educ* 2013;23:628–36.
- 119 Cherry MG, Fletcher I, O’Sullivan H, Dornan T. Emotional intelligence in medical education: a critical review. *Med Educ* 2014;48 (5):468–78.
- 120 Lievens F, Ones DS, Dilchert S. Personality scale validities increase throughout medical school. *J Appl Psychol* 2009;94:1514–35.
- 121 Jerant A, Griffin E, Rainwater J, Henderson M, Sousa F, Bertakis KD, Fenton JJ, Franks P. Does applicant personality influence multiple mini interview performance and medical school acceptance offers? *Acad Med* 2012;87:1250–9.
- 122 Powis DA, Rolfe I. Selection and performance of medical students at Newcastle, New South Wales. *Educ Health* 1998;11:15–23.
- 123 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. *Med Teach* 2005;27:499–503.

- 124 Donnon T, Oddone-Paolucci E, Violato C. A predictive validity study of medical judgement vignettes to assess students' non-cognitive attributes: a 3-year prospective longitudinal study. *Med Teach* 2009;31:e148–55.
- 125 Donnon T, Paolucci EO. A generalisability study of the medical judgement vignettes interview to assess students' noncognitive attributes for medical school. *BMC Med Educ* 2008;8:58.
- 126 Elam CL, Studts JL, Johnson MMS. Prediction of medical school performance: use of admission interview report narratives. *Teach Learn Med* 1997;9:181–5.
- 127 Kleshinski J, Shriner C, Khuder SA. The use of professionalism scenarios in the medical school interview process: faculty and interviewee perceptions. *Med Educ Online* 2008;13:2.
- 128 Patrick LE, Altmaier EM, Kuperman S, Ugolini K. A structured interview for medical school admission, phase 1: initial procedures and results. *Acad Med* 2001;76:66–71.
- 129 Rahbar MH, Vellani C, Sajan 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. *Med Educ* 2001;35 (4):374–80.
- 130 Van Susteren TJ, Suter E, Romrell LJ, Lanier L, Hatch RL. Do interviews really play an important role in the medical school selection decision? *Teach Learn Med* 1999;11:66–74.
- 131 Basco WT Jr, Gilbert GE, Chessman AW, Blue AV. The ability of a medical school admission process to predict clinical performance and patients' satisfaction. *Acad Med* 2000;75:743–7.
- 132 Basco WT, Lancaster C, Carey ME, Gilbert GE, Blue AV. The medical school applicant interview predicts performance on a fourth-year clinical practice examination. *Pediatr Res* 2004;55 (4):350.
- 133 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. *Adv Health Sci Educ Theory Pract* 2008;13:151–62.
- 134 Fan AP, Tsai TC, Su TP, Kosik RO, Morisky DE, Chen CH, Shih WJ, Lee CH. A longitudinal study of the impact of interviews on medical school admissions in Taiwan. *Eval Health Prof* 2010;33:140–63.
- 135 Kreiter C, Yin P, Solow C, Brennan R. Investigating the reliability of the medical school admissions interview. *Adv Health Sci Educ Theory Pract* 2004;9:147–59.
- 136 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. *Med Educ Online* 2005;10:14.
- 137 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. *Med Teach* 2014;36:640–2.
- 138 Eva KW, Macala C. Multiple mini-interview test characteristics: 'tis better to ask candidates to recall than to imagine. *Med Educ* 2014;48 (6):604–13.

- 139 Dore KL, Kreuger S, Ladhani M et al. The reliability and acceptability of the multiple mini-interview as a selection instrument for postgraduate admissions. *Acad Med* 2010;85 (Suppl):60–3.
- 140 Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Med Educ* 2004;38 (3):314–26.
- 141 Hofmeister M, Lockyer J, Crutcher R. The acceptability of the multiple mini interview for resident selection. *Fam Med* 2008;40:734–40.
- 142 O'Brien A, Harvey J, Shannon M, Lewis K, Valencia O. A comparison of multiple mini-interviews and structured interviews in a UK setting. *Med Teach* 2011;33:397–402.
- 143 Roberts C, Walton M, Rothnie I, Crossley J, Lyon P, Kumar K, Tiller D. Factors affecting the utility of the multiple mini-interview in selecting candidates for graduate-entry medical school. *Med Educ* 2008;42 (4):396–404.
- 144 Campagna-Vaillancourt M, Manoukian J, Razack S, Nguyen LH. Acceptability and reliability of multiple mini interviews for admission to otolaryngology residency. *Laryngoscope* 2014;124:91–6.
- 145 Knorr M, Hissbach J. Multiple mini-interviews: same concept, different approaches. *Med Educ* 2014;48 (12):1157–75.
- 146 Uijtdehaage S, Parker N. Enhancing the reliability of the multiple mini-interview for selecting prospective health care leaders. *Acad Med* 2011;86:1032–9.
- 147 Hissbach JC, Sehner S, Harendza S, Hampe W. Cutting costs of multiple mini-interviews – changes in reliability and efficiency of the Hamburg medical school admission test between two applications. *BMC Med Educ* 2014;14:54.
- 148 Eva KW, Reiter HI, Rosenfeld J, Norman GR. The ability of the multiple mini-interview to predict preclerkship performance in medical school. *Acad Med* 2004;79 (Suppl):40–2.
- 149 Eva KW, Reiter HI, Trinh K, Wasi P, Rosenfeld J, Norman GR. Predictive validity of the multiple mini-interview for selecting medical trainees. *Med Educ* 2009;43 (8):767–75.
- 150 Hofmeister M, Lockyer J, Crutcher R. The multiple mini-interview for selection of international medical graduates into family medicine residency education. *Med Educ* 2009;43 (6):573–9.
- 151 Reiter HI, Eva KW, Rosenfeld J, Norman GR. Multiple mini-interviews predict clerkship and licensing examination performance. *Med Educ* 2007;41 (4):378–84.
- 152 Rosenfeld JM, Reiter HI, Trinh K, Eva KW. A cost efficiency comparison between the multiple mini interview and traditional admissions interviews. *Adv Health Sci Educ Theory Pract* 2008;13:43–58.
- 153 Hopson LR, Burkhardt JC, Stansfield RB, Vohra T, Turner-Lawrence D, Losman ED. The multiple mini interview for emergency medicine resident selection. *J Emerg Med* 2014;46:537–43.
- 154 Pau A, Jeevaratnam K, Chen YS, Fall AA, Khoo C, Nadarajah VD. The Multiple Mini-Interview (MMI) for student selection in health professions training – a systematic review. *Med Teach* 2013;35:1027–41.

- 155 Axelson R, Kreiter C, Ferguson K, Solow C, Huebner K. Medical school preadmission interviews: are structured interviews more reliable than unstructured interviews? *Teach Learn Med* 2010;22:241–5.
- 156 Kumar K, Roberts C, Rothnie I, du Fresne C, Walton M. Experiences of the multiple mini-interview: a qualitative analysis. *Med Educ* 2009;43 (4):360–7.
- 157 Quintero AJ, Segal LS, King TS, Black KP. The personal interview: assessing the potential for personality similarity to bias the selection of orthopaedic residents. *Acad Med* 2009;84:1364–72.
- 158 Razack S, Faremo S, Drolet F, Snell L, Wiseman J, Pickering J. Multiple mini-interviews versus traditional interviews: stakeholder acceptability comparison. *Med Educ* 2009;43 (10):993–1000.
- 159 McManus IC, Richards P, Winder BC. Do UK medical school applicants prefer interviewing to non-interviewing schools? *Adv Health Sci Educ Theory Pract* 1999;4:155–65.
- 160 Dowell J, Lynch B, Till H, Kumwenda B, Husbands A. The multiple mini-interview in the UK context: 3 years of experience at Dundee. *Med Teach* 2012;34:297–304.
- 161 Humphrey S, Dowson S, Wall D, Diwakar V, Goodyear HM. Multiple mini-interviews: opinions of candidates and interviewers. *Med Educ* 2008;42 (2):207–13.
- 162 Sebok SS, Luu K, Klinger DA. Psychometric properties of the multiple mini-interview used for medical admissions: findings from generalisability and Rasch analyses. *Adv Health Sci Educ Theory Pract* 2014;19:71–84.
- 163 Dodson M, Crotty B, Prideaux D, Carne R, Ward A, de Leeuw E. The multiple mini-interview: how long is long enough? *Med Educ* 2009;43 (2):168–74.
- 164 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. *Med Educ* 2013;47 (8):801–10.
- 165 Gafni N, Moshinsky A, Eisenberg O, Zeigler D, Ziv A. Reliability estimates: behavioural stations and questionnaires in medical school admissions. *Med Educ* 2012;46 (3):277–88.
- 166 ten Cate O, Smal K. Educational assessment centre techniques for entrance selection in medical school. *Acad Med* 2002;77:737.
- 167 Ziv A, Rubin O, Moshinsky A, Gafni N, Kotler M, Dagan Y, Lichtenberg D, Mekori YA, Mittelman M. MOR: a simulation-based assessment centre for evaluating the personal and interpersonal qualities of medical school candidates. *Med Educ* 2008;42 (10):991–8.
- 168 Gale TC, Roberts MJ, Sice PJ, Langton JA, Patterson FC, Carr AS, Anderson IR, Lam WH, Davies PR. Predictive validity of a selection centre testing nontechnical skills for recruitment to training in anaesthesia. *Br J Anaesth* 2010;105:603–9.
- 169 Randall R, Davies H, Patterson F, Farrell K. Selecting doctors for postgraduate training in paediatrics using a competency based assessment centre. *Arch Disease Child* 2006;91:444–8.

- 170 Randall R, Stewart P, Farrell K, Patterson F. Using an assessment centre to select doctors for postgraduate training in obstetrics and gynaecology. *Obstet Gynecol* 2006;8:257–62.
- 171 Roberts MJ, Gale TC, Sice PJ, 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;68:591–9.
- 172 Ackerman PL. A theory of adult intellectual development: process, personality, interests, and knowledge. *Intelligence* 1996;22:227–57.
- 173 Ackerman PL, Heggstad ED. Intelligence, personality, and interests: evidence for overlapping traits. *Psychol Bull* 1997;121:219–45.
- 174 Patterson F, Ashworth V. Situational judgement tests in medical education and training: Research, theory and practice: AMEE Guide No. 100. *Med Teach* 2015:1–15.
- 175 Powis D. Selecting medical students: an unresolved challenge. *Med Teach* 2015;37:252–60.
- 176 Patterson F, Lievens F, Kerrin M, Munro N, Irish B. The predictive validity of selection for entry into postgraduate training in general practice: evidence from three longitudinal studies. *Br J Gen Pract* 2013;63:e734–41.
- 177 Patterson F, Ferguson E, Thomas S. Using job analysis to identify core and specific competencies: implications for selection and recruitment. *Med Educ* 2008;42 (12):1195–204.
- 178 Stemig MS, Sackett PR, Lievens F. Effects of organisationally endorsed coaching on performance and validity of situational judgement tests. *Int J Select Assess* 2015;23:174–81.
- 179 Griffin B, Hu W. The interaction of socio-economic status and gender in widening participation in medicine. *Med Educ* 2015;49 (1):103–13.
- 180 Medical Schools Council. *Selecting for Excellence: Final Report*. London: MSC 2014.
- 181 O’Neill L, Vonsild MC, Wallstedt B, Dornan T. Admission criteria and diversity in medical school. *Med Educ* 2013;47 (6):557–61.
- 182 Work Psychology Group. *Situational judgement test. UKCAT Technical Report 2013*. Derby: WPG 2014
- 183 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. *BMJ* 2012;354:e1805.
- 184 Patterson F, Prescott-Clements L, Zibarras L, Edwards H, Kerrin M, Cousans F. Recruiting for values in healthcare: a preliminary review of the evidence. *Adv Health Sci Educ Theory Pract* 2015:1–23.
- 185 Martin G, Beech N, MacIntosh R, Bushfield S. Potential challenges facing distributed leadership in health care: evidence from the UK National Health Service. *Sociol Health Illn* 2015;37:14–29.

186 Patterson F, Kerrin M, Gatto-Roissard G, Coan P. *Everyday Innovation: How to Enhance Innovative Working in Employees and Organisations*. London: NESTA, National Endowment for Science 2009.

187 Girotti JA, Park YS, Tekian A. Ensuring a fair and equitable selection of students to serve society's health care needs. *Med Educ* 2015;49 (1):84–92.

188 Nicholson S, Cleland JA. Reframing research of widening participation in medical education: using theory to inform practice. In: Cleland JA, Durning SJ, eds. *Researching Medical Education*. London: Wiley- Blackwell 2015;231–43.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Details of the 194 articles included in the present review.

Table S1. Criteria used to apply the literature search.

Table S2. Numbers of papers identified according to selection methods and research questions.

Table S3. Summary of relevant review findings.