BMA Board of Medical Education

Selection for specialty training

November 2006

improving | health

BMA

Editorial board

A report from BMA Science & Education and the Board of Medical Education

Chair, Board of Medical Education	Professor Sir Charles George
Director of Professional Activities	Professor Vivienne Nathanson
Head of Science and Education	Dr Caroline Seddon
Project manager	Simon Young
Research and writing	EeMun Chen
Editorial secretariat	Nicholas Emery Luke Garland Sarah How Gemma Jackson Joanna Rankin George Roycroft

Cover photograph: Getty Creative Images

Produced by BMA Marketing & Publications

© British Medical Association 2006 – all rights reserved. No part of this publication may be reproduced, stored in a retrievable system or transmitted in any form or by any other means that be electrical, mechanical, photocopying, recording or otherwise, without the prior permission in writing of the British Medical Association.

Board of Medical Education

This report was prepared under the auspices of the Board of Medical Education of the British Medical Association, whose membership for 2005/06 was as follows:

Dame Deirdre J Hine	President, BMA
Dr Michael Wilks	Chair, BMA Representative Body
Mr James N Johnson	Chair, BMA Council
Dr D Pickersgill	Treasurer, BMA
Sir Charles George	Chair, Board of Medical Education
Dr R Hooke	Deputy Chair, Board of Medical Education
Dr R Arnold	
Professor B Avery	
Dr J Bamrah	
Dr S Calvert	
Ms D Covey	
Dr G Dilliway	
Dr D Goodsman	
Professor J Hayden	
Dr L Kapila	
Dr G Lewis	
Professor S Lingam	
Mr M Mak	
Dr H McKeown	
Dr J Oxley	
Dr D Sinclair	
Professor F Smith	
Professor T Weetman	
Dr B Qureshi	(Deputy)
Dr G Buckley	(Observer)

Approval for publication as a BMA policy report was recommended by BMA Board of Professional Activities on 21 November 2006.

Declaration of interest

There were no competing interests with anyone involved in the research and writing of this report. For further information about the editorial secretariat or Board members please contact the Science and Education Department which holds a record of all declaration of interests: info.science@bma.org.uk

Acknowledgements

The association is very grateful for the help provided by the BMA committees and several outside experts and organisations. We would particularly like to thank:

Professor Fiona Patterson, **Professor of Organisational Psychology, City University, London**

Contents

Executive summary 5

Introduction 7

PART 1

- Selection: the challenges 9
- Selection for medical school 9
- Selection for specialty training 12
  - Job analysis and competency modelling 12
  - Selecting for potential 13
  - Selecting out 14
  - Trainees also make choices 15
- Diversity 16
- Summary 17

PART 2

- Selection: the tools 19
  - Application forms 19
  - Biographical data measures 19
  - General mental ability and aptitude testing 20
  - Interests 20
  - Interview 21
  - Knowledge tests 22
  - Personality testing 23
  - References 24
  - Skill/ability tests 25
  - Work sample 26
  - Combining methods and assessment centres 26

Is testing valid and reliable? 28

Conclusion 31

References 32

Executive summary

The restructuring of medical education through Modernising Medical Careers has introduced the foundation training programme, for newly graduated doctors, and run-through specialty training. 2007 will see the first cohort of doctors completing the full foundation training programme and entering specialty training, for which a new selection process is being developed.

Run-through training means the selection point for entry moves from a situation where doctors would have had several years' training and experience in a specialty, to a point where all doctors are deemed to have reached the same level of competence with no discriminating difference between them – the outcome of foundation training.

The report considers selection for specialty training in two parts. Part 1 highlights the challenges faced in producing a selection system for entry to specialty training, including:
- selection prior to entry to specialty training
- selecting potential
- selecting out
- the choices that trainees make
- diversity.

Part 2 provides an overview of different selection tools, what they can do and the relative merits of their use in selection for specialty training, including:
- application forms, interests and references
- interviews
- biographical data
- mental ability and aptitude tests
- skill and knowledge tests
- personality tests.

This BMA report is intended to inform doctors, regulators and education deliverers – particularly those charged with setting standards for entry and those designing the selection system – in order that a selection system is produced based on fairness and flexibility and on clear, uniform and UK-wide principles.

Introduction

In determining a career path one relies on factors such as aspiration, and an understanding of abilities and character. Career pathways in medicine, however, are complex not just because of the rich variety of specialties involved but also because there is an interplay of factors including educational experience, employer's requirements, patient expectation and interaction, and personality attributes. The challenge for education is to equip those in training with the right skills and experience to be able to perform successfully as a doctor now and in the future.

The Postgraduate Medical Education Training Board (PMETB) is ultimately responsible for specialist training and approves curricula for training programmes. A network of regional postgraduate deans administers and monitors the system. In January 2006 PMETB launched its principles for entry to specialist training.[1] These are to inform the development of fair and open processes for the selection of candidates applying for places on specialist and general practice training programmes. It states that 'the selection process must be fair to all candidates who may apply, whether UK, European Economic Area or international medical graduates'. It also states that it is mandatory that 'candidates must be able to demonstrate the competences required at the end of the Foundation programme either by successfully completing that programme or by demonstrating that they have gained those competences in another way', and that 'other evidence that may be sought or presented as part of the selection process may include evidence of excellence in terms of attributes such as motivation, career commitment etc, but no requirement of the completion of a particular post'.

To date, selection to specialty training has followed guidance produced by the National Health Service (NHS) Executive[2] on behalf of the four nations which advocates the use of an appointment committee, application forms, shortlisting based on common, agreed and relevant criteria, a structured interview and a minimum of two professional references. It makes no mention of other selection tools that may be used. Other methods have been introduced but there is no standard selection process across specialties. These other methods include objective-structured-clinical-examination based selection, structured interviews, competency-based questions, assessment centres and electronic application forms; and also the more traditional methods of academic qualifications, curriculum vitae and references. There is great competition for places in the specialist stages of training and, given this competition, it is in the best interests of the profession to ensure that a comprehensive selection system is in place. A good selection system not only enables greater differentiation between candidates through valid, reliable and fair means, but also the identification of individuals who will serve the specialty and patients well.

Modernising Medical Careers (MMC),[3] a Department of Health and NHS initiative, aims to improve patient care by delivering a modernised and focused career structure for doctors through a major reform of postgraduate education. Along with the introduction of foundation programme training,[a] the significant change is run-through training. Traditionally the hurdle to specialty training has been at the point of transition from senior house officer (SHO) to specialist registrar (SpR). In this case, applicants for SpR training had experience of training at SHO level in the specialty in which they wished to continue. With run-through training the point of selection for specialty training, or general practice, is now at the transition from foundation training to the equivalent of SHO-level training. At this point in their medical education, applicants will not necessarily have had the opportunity to receive exposure to the specialty in which they plan to practise. MMC also plans to introduce a formalised training programme for academic medicine and, so, there will be another pathway for which the selection process needs careful consideration.

---

[a] Foundation programme. A two-year structured educational programme which aims to provide continuity from medical school through to specialist training. Foundation year one is equivalent to pre-registration house officer while foundation year two is equivalent to first year senior house officer. [4]

This BMA report builds on the commitment to ensure that entry to academic, specialist and general practice training is on a fair, flexible and competitive basis, and on clear, uniform and UK-wide principles. The report is divided into two parts. Part 1 acknowledges and discusses the questions that designing and implementing a selection process pose and part 2 provides an overview of the variety of selection methods that are available and used.

Part 1

Selection: the challenges

Quality healthcare relies not only on technology and advances in medicine but, most importantly, on the people who deliver such care. To ensure quality doctors there must be an efficient and effective selection process for entry into training. Through the identification of necessary competencies or knowledge, skills, abilities and attitudes, medical educators attempt to select individuals that either meet, or have the potential to meet, these ideals. People are the most valuable resource of any organisation, healthcare organisation or otherwise, but they are also very costly and the most unpredictable of resources. In selecting star performers, medical educators and medical organisations mitigate some of this cost and unpredictability. Recruitment and selection system design needs to consider its outcomes – quality healthcare – and a multitude of factors. These include:

- complementing existing selection practices both at undergraduate and postgraduate education and beyond
- the academic institution's strategies and image
- patient needs
- sources of students
- advertising media and content
- legislation
- selection criteria
- testing, reliability and validity of selection methods
- interviews and selection decisions.

Compounding these many factors is the urgent necessity to revisit selection into specialty training. The recent changes to medical education training mean that in 2007 the first cohort who have completed foundation years one and two will enter specialty training. The General Medical Council's (GMC) guidance on undergraduate medical education, *Tomorrow's Doctors*,[5] heralded a change from emphasis on the acquisition of knowledge to 'a learning process that includes the ability to evaluate data as well as to develop skills to interact with patients and colleagues'. Medical education is described by the GMC as a continuum, so in postgraduate years students are building on attitudes, knowledge, understanding and skills acquired at an undergraduate level. Further, MMC aims to develop demonstrably competent doctors who are skilled communicators and effective team members.

This shift from time-based/knowledge-based training to competence begs the question of how this competence will be assessed and how we ensure that potential doctors have the right competencies (either acquired or inherent) for the right specialty. Not only does this require high quality training at both undergraduate and postgraduate levels but that the selection processes to enter education at different points are sufficiently robust. Thus, competent doctors stay within the profession and that those who may bring the profession into disrepute or are low performers are excluded or do not progress unless they respond satisfactorily to remedial training.

Selection for medical school

'Although academic achievement is only the qualifying standard for entering the real field of selection, like the Olympic qualifying standard is to selection for the national team, it is overwhelmingly the strongest element in selection. Unlike all the other desirable attributes of personality, attitude and interest, examination results look deceptively objective. Relatively objective they may be but they are still poor indicators of the potential to become "a good doctor" – a product difficult to define, not least because medicine is such a wide career that there may be many different sorts of good doctors – but they all need the appropriate knowledge, skills, and attitudes for effective medical practice and the ability to use them competently.'[6]

Selection processes occur much earlier in medical education than at the transition to specialist training. As such, it is useful to discuss the assessment and selection systems that are in place when one enters medical school, at undergraduate level, as they may impact on selection procedures later.

There is currently little guidance available on the development, implementation and evaluation of selection processes. For example, in the report by Schwartz,[7] the steering group notes that, in relation to medical school admission, the current evidence base for relevant, reliable and valid measures shown to predict undergraduate success is small. The steering group recommended the establishment of a central specialist source of advice for institutions on admissions, which could play a significant role in developing common agreement across the sector on the reliability and validity of methods of assessment.

Selection to medical school in the UK is generally on the basis of, or a mix of, factors such as academic record, personal statements, references and/or admissions tests. A summary of the selection methods is given in the boxes below.

Academic records

All medical schools set a minimum standard of at least ABB at A-level but the typical grades required at first attempt for entry to medicine vary by medical school.[6]

Examination results may be problematic as a selection tool because, for example, they do not take account of late developers, differences in educational opportunity, support for home study or the effects of ill health on the day. Nevertheless, A-levels are predominantly relied on. McManus, Powis & Wakeford et al[8] note that there are widely and strongly held beliefs about the efficacy of A-levels in predicting medical school outcomes, that is, that they do not have any predictive validity. Instead, their review of literature finds that A-levels have a clear relationship with degree level at medical school, and in general university studies. However, the practical application of A-levels for admission has become increasingly problematic due to the numbers of medical school entrants who are achieving high grades, and therefore A-levels have low discriminative ability.

Aptitude tests

A number of medical schools in the UK, and internationally, have introduced tests on admission to medical school. These are usually adopted in conjunction with other methods, for example, combined with an interview, academic record and personal statement in order to aid in the selection of applicants. Two tests used in the UK are the UK Clinical Aptitude Test (UKCAT) and the BioMedical Admissions Test (BMAT).

The **UKCAT**[9] was used by the majority of medical schools for the first time during summer 2006 as part of their process for selecting medical students. It consists of four subtests which assess verbal reasoning, numerical reasoning, abstract reasoning and problem solving. There will be a more extensive non-cognitive element from 2007 onwards which is envisaged to look at attributes such as empathy, conscientiousness, altruism, integrity, resilience and emotional stability.

The **BMAT**[10] is a subject-specific admissions test taken by applicants to Cambridge, Imperial College London, Oxford and University College London medical schools. The BMAT was introduced by these medical schools to help differentiate between large numbers of well qualified applicants of which the great majority have, or are predicted to get, three A-levels at grade A. This aptitude test assesses ‘thinking skills’, and is designed not only to help choose between well qualified candidates, but also to provide a way of assessing the potential of students whose ability might not be reflected in their grades.

Interviews

The majority of medical schools use an interview as part of their selection strategy.[6] The body of research on interviews favours structured interviews over unstructured interviews. Structured interviews tend to have substantial predictive validity with questions based on the identified tasks, behaviours, knowledge, skills and abilities of the job.[11,12] Unstructured interviews have no fixed format or set of questions to be answered. The interviewer may ask different questions of each candidate, with responses which are not usually scored leaving an overall evaluation based on summary impressions and judgements.

References and personal statements

References and personal statements are typically used by medical schools but have shown to have variable predictive validity. A pilot study of these assessment techniques found that the amount of information in a reference did not reliably predict performance in medical school, although information in personal statements appeared to relate to grades achieved in clinical years.[13]

The application requirements for a medical degree are not standardised and different medical schools have different requirements for entry and different selection systems. While the selection methods are a useful aid in selecting between candidates in terms of who is likely to successfully complete a medical degree course, they are not a predictor of whether or not someone will make a good doctor.

Selection for specialty training

There are a number of factors one may need to consider in selecting doctors for entry to specialty training. Is there overarching medical education, work experience, circumstances, physical make-up, abilities, personality and motivation one would expect of all doctors, or is there a different one for each of the specialties and subspecialties? Organisational fit is also a consideration. Differences in working and team environment can impact on an individual's suitability for different roles. For example, a decreased formality in a district hospital or a hospital with an academic as opposed to a clinical emphasis may suit different individuals.[14]

Following completion of undergraduate medical education, newly-qualified doctors undertake a two-year foundation training programme. The foundation programme aims to produce doctors who are both competent at dealing with the acutely ill and who are effective at communicating with patients and colleagues.[3] Successful completion of foundation training requires doctors to show that they have achieved a number of competencies. These competencies are assessed using four different assessment exercises,[b] the outcomes of which form a profile and a development plan for each doctor based on their strengths and any areas identified for development.

By the time a doctor has undertaken and successfully completed undergraduate education and foundation training they have been assessed to ensure that they have the academic ability to complete a demanding medical degree course and are effective communicators with patients and colleagues. So, in relation to selection for specialty training, the following questions might be considered:

- do, or should, these qualities need to be tested again?
- if all doctors have attained this level, how do we discriminate between candidates at the selection point for specialty training?
- what should be assessed for selection to specialty training?

With the outcome of foundation training being that all successful trainees will have gained the same competencies, selection processes will have the dilemma of selecting trainees for specialties who are already deemed to have reached the same level of competence.

Job analysis and competency modelling

Some specialties have made attempts to determine what behaviour and characteristics would constitute a doctor suitable to that specialty. This process is referred to in the industrial and organisational psychology, and human resources literature, as job analysis. Job analysis is a process to identify the tasks, behaviours, knowledge, skills and abilities that embody a job. These may be identified through a number of methods, for example, direct observation, interviews, questionnaires,[c] critical incident technique,[d] repertory grid technique,[e] daily diaries, task checklists and document review. Each method has varying levels of efficiency and effectiveness and trade-offs may need to be made based on resources and level of depth needed. Various techniques may be used for incremental validity and triangulation (cross-validation). Job analysis techniques are used in Agenda for Change,[15] a pay system for all directly employed NHS staff except doctors, dentists and very senior managers.

---

[b] Further information on these tests can be found on the MMC website www.mmc.nhs.uk/pages/assessment

[c] Some well-known general questionnaires are the Position Analysis Questionnaire, Functional Job Analysis, Job Element Inventory, Threshold Trait Analysis and the Ability Requirements Scale.

[d] Critical incidence technique. This technique asks subject matter experts to describe incidents which are examples of excellent/poor performance and describe the incidents leading up to the event, the knowledge, skills and abilities associated with the incident and why it was effective/ineffective.

[e] Repertory grid. An interviewing technique devised by George Kelly to complement his Theory of Personal Constructs. It uses a matrix to elicit how individuals think about concepts according to its particular attributes. In this case the good doctor could be contrasted with a poor doctor by the characteristics and aspects of their performance.

Selecting respondents to help in the job analysis process can be problematic. Job analysis typically involves subject matter experts such as incumbent doctors and supervisors. A number of factors can influence the information gathered, such as years of experience and motivation to make the job sound more important.

Once tasks are identified these are generally clustered into primary roles and/or functions and requisite behaviours. These roles and behaviours can then be further analysed to identify the competencies needed to perform them. These competencies, and their definitions, then form part of a person specification and key selection criteria, which can help focus interviewing, screening, short-listing and test selection. The outcomes of job analysis can also be incorporated into performance appraisal and training and development processes.

Selecting for potential

While an individual may be able to complete medical school, specialty training and become a consultant, does it follow that they have the skills and attitudes to excel as a doctor in practice? The study of competencies for general practice regarded some desired attributes as more suitable for training, rather than assessed in selection. Some competencies and behaviour can be learned and/or rehabilitated; however, at this stage, it is unclear which ones these are. For example, manual dexterity in the case of surgery could be considered to be a skill that someone either does or does not have. On the other hand it has been argued that manual dexterity is an ability that may be improved through surgical practice.[16, 17]

Arguably, training may be more important than selection.[18] Perhaps all that selection requires is identifying minimum requirements and then it is the training process which defines and ensures a competent doctor.

The GMC's *The new doctor*[19] explains the outcomes doctors must achieve before they are granted full registration. These seven outcomes, in line with *Good Medical Practice*, [20] are:

- good clinical care – doctors must practise good standards of clinical care, practise within the limits of their ability, and make sure that patients are not put at unnecessary risk
- maintaining good medical practice – doctors must keep up to date with developments in their field and maintain their skills
- relationships with patients – doctors must be able to develop, encourage and maintain successful relationships with their patients
- working with colleagues – doctors must work effectively with colleagues in medicine, other healthcare professions and allied healthcare workers
- teaching and training – doctors have teaching responsibilities to colleagues, and must develop the skills, attitudes and practices of a competent teacher
- probity – doctors must be honest
- health – doctors must take appropriate action to maintain their health and protect patients.

In addition to these principles of professional practice, trainees must develop specific skills, experience and knowledge which must be included in all training programmes.[20] These are:

- the scientific basis of practice and treatment
- diagnosis and treatment
- clinical and procedural skills
- communication skills
- teaching and learning skills
- personal and professional skills
- the changing patterns of healthcare
- legal and ethical issues
- disability and rehabilitation
- the health of the public
- the individual in today's society.

Therefore, quality doctors should theoretically be produced through both robust selection and training procedures. Although, as stated above, it is unclear what is more appropriately assessed at selection and what can be trained. Other than relying on professional judgement, how does one select for potential?

Selecting out

Instead of struggling to identify who to select, perhaps another approach should be taken, that is, to 'select out'. There are traits and behaviours that the profession would consider to be that of 'a bad doctor' and individuals who possess these could be excluded from training, for example, psychopathy and antisocial personality.[21]

Dame Janet Smith, chair of the public inquiry into the actions of serial killer GP Harold Shipman, proposed in a GMC conference that an ethics test be administered to determine who should be allowed to practise as a doctor.[22] Moral orientation as a predictor of moral decisions of medical students has been studied by Bore, Munro & Kerridge et al.[23] They developed a test of moral orientation, Mojac, which 7,864 medical school applicants and students in Australia, Israel, Fiji, New Zealand, Scotland (*n*=508) and England (*n*=69) completed. The test involves reading hypothetical dilemmas, and participants are asked for their level of agreement with a series of statements relating to the needs of individuals or to the needs and norms of society. Participants must also make a forced choice final decision for each dilemma. This kind of test could be used to select out those who appear to have an extreme moral orientation. Further research is needed though as the predictive validity of this particular test has not yet been established.

Selecting out those who engage in risky behaviour may also be advantageous. It could be argued that risk perception and risky behaviour can be considered both as positive and negative, for example, taking extra risks may be a necessity of particular fields of medicine. Firth-Cozens, Cording & Ginsburg[24] consider that low risk perception, as well as sensation seeking, Type A personality,[f] high self esteem and psychological ill health are individual characteristics that would be most likely to affect an individual's regard for patient safety. While selecting for safety is attractive, further research is required before it can be instituted in any selection process.

---

[f] Type A personality is characterised by aggressiveness, competitiveness, impatience and tenseness. It is considered to be related to stress-related symptoms and increased risk of coronary heart disease.

There has been some research in the USA on the association between disciplinary action by medical boards and a documented lack of professional behaviour when those physicians were medical students.[25, 26] Papadakis, Teherani & Banach et al[26] compared the academic records of 235 medical school graduates from three medical schools who were disciplined by state medical boards with a matched control group of medical graduates. They found that disciplinary action was associated with unprofessional behaviour in medical school (such as severe irresponsibility and severely diminished capacity for self-improvement). There were some links between low scores on the US Medical College Admission Test, and low grades in the first two years of medical school and unprofessional behaviour.

Trainees also make choices

While specialties may be able to make choices about who they admit into training, it must be acknowledged that trainees self-select into specialties. Choices may be made on factors other than purely aptitude and competency, such as:

- domestic circumstances
- flexibility
- working conditions/hours
- financial circumstances while training
- promotion prospects
- eventual financial prospects
- anticipated ease of obtaining a post
- organisation of training programme
- self appraisal
- advice from others
- inclination before medical school
- student experience
- experience of jobs in training
- availability of information about specialties
- availability of posts in preferred region
- enthusiasm/commitment
- influence of family members and/or influence of a consultant.[27]

McParland, Noble & Livingston et al[28] specifically investigated student attitudes toward psychiatry as a studying and career preference. They found that positive attitudes and intention to pursue psychiatry as a career were influenced by encouragement from senior psychiatrists, direct involvement in patient care, seeing patients respond well to treatment and influence and encouragement by someone. Their conclusion was that undergraduate teachers may have a role to play in influencing the choice of one's specialty. So those who eventually enter specialty training are a result of a complicated interplay between the individual, medical education and the selection process.

Testing can also be used to give career guidance, for example, the Sci45[g] Specialty Choice Inventory[29] matches career choices with the profile of the trainee (based on skills, attributes, aptitudes and lifestyle aspirations). The test consists of 130 items and matches according to 46 specialties, including general practice. The concept of links between one's attributes and medical career has also been illustrated.[30] They found that medical career preferences may be separated into six general areas or specialties: surgery, hospital medicine, psychiatry, public health, administrative medicine and laboratory medicine, and suggest that interest, ability and personality are factors in determining preference. The application of the Myers-Briggs Type Indicator, a personality test usually used for career and team development, to US medical school graduates has shown how differences in personality can relate to specialty choice.[31]

---

[g] Sci45 was updated to Sci59 on 1 September 2006. Further information can be found at iet.open.ac.uk/about/oucem/index.cfm?id=5835#SCI45 (accessed November 2006)

Diversity

The questions and discussion above hinge on the outcome we want from a selection system. There are a number of possible outcomes, for example, finding individuals with the ability to learn, those who are already accomplished, those who are and will be high performers, have the requisite skills, have the right attitudes, selecting out underperformers and have the motivation and will to work hard. There is a danger that selection based on a number of criteria will result in doctors who only fit that model, and could bring about a workforce that does not have the diversity of the population it is there to serve.

There is a need to distinguish between the large numbers of candidates for specialty training, and selection on certain characteristics may be the way to achieve this. What also needs to be addressed is widening participation in medical school and specialty training for groups who have been traditionally under-represented in medical education. Educational policies will also influence levels of participation, for example, the government's target of 50 per cent of young people entering higher education by 2010.[32]

The profession should reflect the population it serves. In 2005, 60 per cent of accepted applications from UK-based students to medical and dental school came from families with managerial and professional backgrounds. 32 per cent of accepted applications from UK domiciled students came from minority ethnic backgrounds.[33] In 2005, 52 per cent of white UK-domiciled applicants to medical school were accepted compared with 23 per cent of black UK-domiciled applicants and 39% of asian UK-domiciled applicants. Acceptance rates differ between ethnic groups which may in part be due to selection processes.[34] Selection processes must not only be reliable and valid but fair and non-discriminatory on the grounds of age, colour, sex, religion, politics, marital status, sexual orientation, membership or non-membership of trade unions or associations, or ethnic origin. These issues are to be kept in mind at all stages of the selection process as it affects what might be used in terms of tools and tests, and whether interviews should be used and how they should be conducted.

Items in a test may be culture bound, that is, require knowledge which is common to one group but rare in another. A comprehension test that includes proverbs or idioms which are idiosyncratic to one culture would place some groups of test takers at an advantage over others. Equally, cultural factors can impact on scores in personality tests of factors such as introversion and anxiety.[35]

The West Midlands Deanery has operated a centralised GP trainee selection process since 2000. An audit was conducted of placement success as a function of ethnicity and country of qualification.[36] Candidates were required to complete a standard application form, shortlisted based on pre-set criteria from the application form and attend an assessment centre which contained two interview panels and a role-play exercise. Total scores were calculated based on all these methods and a score rank and preference determined placements, with unplaced candidates entering a national clearing scheme. An earlier audit showed that the process may have been favouring some groups over others and changes were made within the different stages to reduce this possibility. Despite these changes the audit found that UK-trained students were statistically significantly more likely to be shortlisted, with African- and Asian-trained least likely to be shortlisted. Furthermore, UK white candidates had significantly higher average total scores for all selection rounds over four years, while those from an Asian background were significantly lower. The results suggested that either discrimination may have occurred or there are actual performance differences between certain groups. The study shows how important it is to monitor performance in selection systems over time and the impact culture, education background and ethnicity can have.

Summary

The conundrum of selection for specialty training is a difficult one. With the introduction of MMC and foundation training, the time at which doctors are selected for specialty training has moved from a situation where doctors might have had from two to five years' training and experience in that specialty, to a point where all doctors are deemed to have reached the same level of competence with no discriminating difference between them.

Some candidates may have been fortunate enough during foundation training to have experienced their chosen specialty first hand. Others may not, yet may still feel that their chosen specialty is appropriate for them and will expect the right to compete on a level playing field for entry to that specialty with their more fortunate counterparts.

The principles set-out by the PMETB,[1] the regulator and standard-setter for postgraduate medical training and education in the UK, enforce this level playing field by stating that 'other evidence [other than that of demonstrating achievement of the foundation programme competencies] that may be sought or presented as part of the selection process may include evidence of excellence in terms of attributes such as motivation, career commitment etc, but no requirement of the completion of a particular post'. In other words, trainees will not have been expected to have completed a training post in the specialty to which they are applying.

Part 2

Selection: the tools

This section provides an overview of different selection tools and what they can do. It looks at the relative merits of their use in selection for specialty training, including the predictive validity of the selection tools, ie how successful the tools are in predicting set criteria.

Application forms

Applications will usually exceed positions available and it is not viable to interview and assess all applicants. Therefore, it is necessary to discriminate between candidates. As with the whole selection process, valid selection criteria should be applied to the applicant pool. Application forms allow the standardisation of applicant information, which is especially useful when there are many applications, as there will be with foundation year two doctors entering specialty training. Ultimately, application forms aid the screening process. A curriculum vitae or supporting information are hard to use to compare applicants due to different formats and information provided. Application forms will typically ask for education and qualifications, leisure and interests and questions specifically related to desired competencies as identified through job analysis.

The NHS and the UK health departments promote and implement equal opportunities policies. Discrimination on the grounds of age, sex, marital status, race, religion, belief, sexual orientation, colour or disability is prohibited. Eliciting such information is unlawful unless these attributes are a genuine and demonstrable requirement for the job/training position. Each stage of the recruitment and selection process must heed this policy, from the advertisement, shortlisting, selection method utilised and selection system implemented.

Competencies that are desired or required for the position may be used to compare applicants. Those that meet the essential criteria, but not necessarily the desirable ones, would proceed to interviewing and/or other selection process. Clear, documented records should be kept throughout the process. The decision-making process should be standardised and based on a scoring system with agreed criteria.

Biographical data measures

Biographical data measures are questions about past life experiences and interests. They may include early life experiences in one's family, at school and hobbies. In some cases application forms incorporate biographical predictors or 'biodata' to produce a weighted application blank. Biodata techniques rely on the principle that past behaviour is the best predictor of future behaviour. Biodata can be developed from job analysis and validated against a high performing population of job incumbents. Analysis suggests that biographical data measures have reasonable predictive validity for job performance and training performance, but no real validity above general mental ability (GMA)/intelligence tests.[11] They suggest that this is due to biographical data measures being, in part, indirect reflections of GMA. Biodata has also been used to predict turnover, absenteeism, delinquency, substance abuse, promotion, achievement and accidents.[37] In addition, it is suggested that application forms using biodata can have low adverse impact for minority ethnic candidates.[37]

Research into the incorporation of biodata into selection systems has been conducted in a number of industries, although medicine does not appear to feature. Nevertheless, the principles of biodata have been borrowed in the development and implementation of application forms for entry into medical school, foundation programmes and specialty training. If these forms are not empirically derived or based on job analysis or competency modelling their predictive validity is limited and could call into question their procedural fairness and necessity.

Weighted biodata application tools can be costly and technically difficult to construct, but they can be effectively administered once constructed. Validated biographical data measures can be purchased commercially and administered and scored by the vendor.

General mental ability and aptitude testing

Aptitude tests generally refer to tests which measure an individual's capacity to learn or acquire a particular knowledge and/or skill, and are usually good predictors of trainability.[38] Kline, in his standard text for organisational and educational psychologists *Handbook of psychological testing*,[35] argues that aptitude is a misnomer as it often refers to both abilities and attainments and, as it generally equates to general intelligence, it is not differential and should be abandoned as a psychometric or scientific concept. Nevertheless, 'aptitude' is often used and in this report will adopt the meaning of 'capacity to learn and acquire particular knowledge and/or skills'. It is acknowledged that tests referred to as 'aptitude tests' are often a mixture of tests of intelligence, ability and attainment. These tests generally include verbal reasoning, numerical reasoning and/or pattern matching, and are timed, multiple choice and pencil-and-paper tests which are now computer-analysed.

In developing aptitude tests for a population of students entering medical school, or trainees entering specialty training, restriction of range and potential ceiling effects should be borne in mind. Restriction of range refers to a sample who are more likely to attain scores that cluster in one section of a normally distributed bell curve. For example, those applying to medical school are likely to have high intelligence and therefore score highly in an aptitude test. In this case, the test may not be able to distinguish between applicants and therefore any reliability and validity data should be treated with caution. Ceiling effects are similar and should be considered when constructing a test. Ceiling effect refers to the inability of a test to adequately assess the performance of high scorers due to a low number of difficult items.

While research suggests that a test of aptitude would be ideal for selection to specialty training, its inclusion at this stage of medical education may be redundant. Many medical schools are, or will be, using aptitude testing as part of the admissions process. If those tests have high reliability there may be no additional value gained in administering another aptitude test at selection to specialty training.

Interests

There has been some research testing the hypothesis that candidates who have interests which match the content of their jobs will have higher job performance than those with non-matching interests. This research finds there is relatively poor correlation. There are indications that interests substantially influence the jobs that individuals prefer and enter. Following job acquisition, however, the quality and level of performance appears to be determined mostly by 'cognitive horsepower' and personality traits like conscientiousness, rather than interests.[11]

Interests are often used by selectors at medical school. In a recent survey medical students felt pressure to bolster their curriculum vitae with extracurricular activities, but they also recognised that outside interests made for a more well-rounded doctor and believed that employers should be taking notice of these extracurricular activities.[39]

Using interests as a selection tool should be viewed with caution, not just due to its low predictive validity but that it may also unfairly discriminate. Some students may suffer lack of access to opportunity as different schools have different links and opportunities. Work observation and work experience is often held in high regard, but access to these is often seen as available only to those with personal connections.

The belief that interests predict performance is evident in commentary on what should be considered when selecting for specialty training today. A recent editorial suggested that clear evidence of a surgical interest, such as award of a surgical prize, a surgical elective, surgical special study modules, a log book with evidence of significant operative exposure and membership of the Student Surgical Society at medical school should be taken into account for selection into surgical training programmes as it indicates commitment and enthusiasm.[40]

Interview

NHS guidance recommends a structured interview for entry into specialty training.[2] Unstructured interviews have less predictive validity, reliability and standardisation than structured interviews, but are still used and preferred by many organisations.[12, 42] Structured interviews place more burden on the instrument and less on an interviewer's interviewing and assessment skills, which is a great source of variability and error in unstructured interviews. Advantages of structured interviews are not only their psychometric properties but also their legal defensibility and more positive applicant and interviewer reactions.[42, 43] Schmidt & Hunter's meta-analysis found that structured interviews have predictive validity for job performance above that of unstructured interviews. Smith & George[44] suggest that the increase in predictive validity with the change in structure is due to the instrument focusing both the interviewer and interview, and increases the likelihood that the interview will be based on job-related information gathering.

Structured interviews require an interview schedule. Questions target competencies identified during job analysis to ensure not only that the necessary information is collected but that selection decisions are legally defensible. Interview schedules may also include behavioural indicators and rating scales. These not only ease note-taking but also allow quantification and consistency. Scales enable an overall competency score to be arrived at and can aid selection decisions. Interviews may include behavioural questions (questions about the past) and/or situational questions (responses to hypothetical situations). Studies comparing interview approaches have found slightly higher validity for the situational interview.[45] Chapman & Zweig[42] suggest that the use, or not, of predetermined and standardised scoring procedures could be influencing predictive validity rather than the interview approach per se.

It is particularly pertinent to be mindful of potential biases and errors when conducting the interview which may affect outcomes. Examples of these are:

- reliance on first impressions
- stereotyping
- judging by appearances
- confirmatory bias[h]
- halo effects[i]
- inconsistency

---

[h] When the interviewer seeks out information that supports or confirms their hypotheses.

[i] A perceived positive feature or trait is extended to other aspects of the person. For example, a good looking person is also perceived as a good job performer.

- similarity effects[j]
- negative information effect[k]
- cultural differences.[12, 46]

The structured interview does much to minimise these errors and biases but it is always important to be aware of the possible effects they may and do have. Panel interviews, multiple interviewers and training of interviewers may also reduce some of these effects. Equal opportunities legislation should also be referred to as interview questions should not be discriminatory and should be solely related to the agreed criteria.

Knowledge tests

In many jobs, knowledge tests are inappropriate because they cannot be used to evaluate and hire inexperienced workers. This may also be the case for selection for specialty training. While trainees are not inexperienced nor lack specific job knowledge, a job knowledge test may prove unnecessary as all trainees should possess a certain level of knowledge at the end of foundation training year two, and have undergone many assessments of knowledge already.

Schmidt & Hunter[11] in their meta-analysis of selection procedures for predicting job performance found that knowledge tests can have high validity relative to other methods. Knowledge tests tend to be used less commercially as they are time consuming and expensive to construct. For the medical profession, there may be substantial practical value in investing in knowledge tests that not only assess knowledge but that predict performance. While in theory knowledge tests at selection may seem attractive, their predictive validity in the context of performance in medical training and doctor performance on the job has not been established. The outcome of a national knowledge test is attractive in that it may be possible to rank specialty training candidates and aid the selection process. This benefit needs to be weighed against the potential costs and challenges.

The use of knowledge tests for selection into general practice training has been investigated in the Netherlands.[47] The study concluded that the selection test used had no predictive validity for performance in an assessment eight months later, nor at certification. While the sample was relatively small and the outcome measure may have influenced results, it does provide some support for the argument against the use of knowledge tests for selection. Knowledge tests are also used in the UK. The London Deanery has a pre-eligibility test of basic medical knowledge consisting of multiple choice questions and matched extended questions. Doctors must pass this test before applying to the London Deanery for general practice training. Those who are eligible are then invited to sit a written paper under exam conditions which consists of seven competency-based essay type questions. With the current structure of assessments in medical education, and without an established reliable and valid knowledge test in place, the use of a knowledge test in selection testing for specialty training appears to be impractical.

---

[j] This is when similarities between the interviewer and applicant influence interview processes and outcomes. In general, research reports that interviewers often rate applicants higher who have similar attributes to themselves. These attributes may be attitudes, race or sex.

[k] Negative information about the interviewee may be weighted more heavily than positive information. The order in which positive and negative information is presented may also affect overall judgments.

Personality testing

The BMA's annual representative meeting in 2002 resolved 'That personality testing is currently not sensitive enough to establish what makes a safe or good doctor. It should hold no place in:

i) medical school selection;
ii) assessment of qualified doctors referred to the National Clinical Assessment Authority'.

This approach is understandable given that many have reservations regarding the use of selection testing more generally, and not just personality testing. The National Clinical Assessment Service (NCAS), a division of the National Patient Safety Agency, now utilises behaviour assessments in the event of doctors whose performance gives cause for concern. The behavioural assessments include personality tests and an interview with a non-clinical behavioural psychologist. NCAS notes that its evidence base is still very small, but it has found patterns in the personal qualities of those assessed.

Personality traits are generally thought to be relatively enduring, although there is the view that while personality changes do occur these tend to stabilise with limited change at adulthood, usually around age 30.[48] Roberts, Walton & Viechtbauer's[49] meta-analysis of longitudinal studies concludes that personality changes across one's life. Contrary to perceptions that personality develops and is in flux in childhood and adolescence, they found that personality traits changed most in young adulthood (between 20 and 40 years old). Extroversion (social dominance), conscientiousness and emotional stability increased in young adulthood, while extroversion (social vitality) and openness to experience increased in adolescence then decreased in old age. Agreeableness does not appear to increase significantly in any particular age period but does appear to increase across a lifespan.

Before an appropriate personality test can be applied it must be clear what the 'doctor personality' is. There has been little examination of what this personality may consist of. Personality tests usually consist of paper-and-pencil or computer-assisted tests which ask the test taker for their agreement with a number of statements or to make preference choices from a list of options. They may measure one personality trait or a number of personality traits.

Outside the medical profession, integrity and conscientiousness tests have been found to be strong predictors of job performance and training performance.[11] Integrity tests are most used to reduce the probability of counterproductive job behaviours, for example, drinking and drug taking, fighting and stealing. Interestingly, not only do integrity tests predict these types of behaviours but they also predict overall performance. Integrity tests are found to measure other personality traits of conscientiousness, agreeableness and emotional stability.[11]

Until there is further evidence of what traits are desirable in a doctor, personality testing could remain underutilised. The difficulty of doctor profiling is compounded by the many specialties which exist under the umbrella term 'doctor', and the question of whether there is an overarching doctor personality or one for each of the specialties and subspecialties. Indeed, longitudinal studies of large cohorts of medical students in the USA found that compared to other groups of students, for example in business, law, physical science, engineering and humanities, medical students were more likely to be evenly distributed between different personality types.[31] With medicine being such a diverse profession, it is argued that all personality types are needed. These studies focused on the associations between personality type and specialty choice, rather than certain personalities being more suited to particular specialties. Other studies have also found differences in personality between specialties, for example, surgeons were more likely than geriatricians to display a preference for a detached, logical and objective decision-making style, rather than valueorientated, subjective and personal.[50] Those who make judgments subjectively and personally and consider how choices affect others were found to choose family medicine significantly more often.[31]

Fitzgerald[51] found that, for surgeons, superior job performance, as rated by fellow surgeons, nurses and managers, was associated with those who were more traditional rather than innovative (less likely to come up with new ideas or deviate from well proven methods), less likely to argue or have strong views and have higher emotional control. General practice has recently added tests of empathy and sensitivity, coping with pressure and professional integrity to their selection processes. While findings are preliminary, there seems to be an association between job performance and higher scores in empathy and coping with pressure.[52]

There appears to be value in personality testing as part of the selection toolkit, but further specific research would be necessary before considering implementing it as part of the selection process for specialty training. Certainly, research into all specialties needs to be completed and job performance outcomes incorporating patient perspectives may be useful.

References

Reference checks are heavily relied on in employment decision making, both in hiring of doctors and in medical education. A recent editorial in *The Annals of the Royal College of Surgeons of England* called for not only the consideration of performance at an undergraduate level, such as surgical prizes and high academic achievement in clinical subjects, but also references from surgical attachments to aid the selection process.[40]

The NHS guidance[2] states that a ‘minimum of two professional references is essential’. The function of references is to verify factual information and comment on the strengths and weaknesses of a candidate as an indicator of that person’s suitability for appointment. As discussed previously, the scoring of a reference may be problematic. Further, the quality of a reference may depend more on the writer than the person that the referee is writing about.

‘If a senior house officer is looking for a specialist registrar job I spend about an hour or so at the word processor; two to three hours if a specialist registrar is looking for a consultant job. Having given trainees a promise to do my best, they trust me to do my best. But why bother? I ask the question since references do not seem to mean what they used to mean, what I think they mean, what trainees think they mean, or what they ought to mean. And they are not read.’[53]

To mitigate the potential drawbacks of references, structured reference forms or structured telephone reference checks may be used. This ensures the same questions are answered by all referees in order to provide a uniform and comprehensive picture of a candidate. Structured telephone reference checks have shown some value in predicting job performance.[54]

Skill/ability tests

While tests of general mental ability (GMA) are shown to be predictive of job performance generally, this does not negate the use of skill or ability tests, or other selection tools, to select for other factors that are important for certain jobs.

Little research has been completed on the links between ability tests and performance as a doctor. There has been some within surgery but other specialties are under-represented. Intuitive links exist between certain abilities and what one would consider to be a good surgeon, for example, operative ability, manual dexterity, ability to convert 3D images to 2D and vice versa, complex problem solving ability and visual spatial ability. While these have been discussed as possible skills to test for at selection there is very little research on whether these skills are predictive of surgical ability.[55]

Bann & Darzi,[56] in relation to surgical training, review some of the literature and find weak, no, or negative correlations between operative ability, manual dexterity and visual spatial ability and performance. It should be noted though that flawed methodology, small sample sizes and instrument reliability and validity may be contributing factors to these results. For example, the tests used were initially constructed for use in factory production line settings and not surgery. The studies reviewed, involve the application of a number of manual dexterity tests to surgeon samples.

Francis, Hanna & Cresswell et al[57] tested 20 'master surgeons' and twenty medical students using spatial and dexterity tests. While the surgeons made significantly fewer errors on the measure of hand-eye coordination, there were no significant differences in scores between the two groups in the test of hand-eye coordination and manual dexterity. In fact, the medical students fared significantly better in a space relations test, which measures ability to convert 2D images to 3D. This highlights the difficulty in applying such tests at selection to specialty training and whether performance in such tests prior to training is predictive of performance once the trainee is on the job.

While not intended for selection, there have been tools developed and implemented for assessing surgical skills. These include the Imperial College Surgical Assessment Device[58] which tracks motion in dexterity tasks and measures the time taken, the number of movements and path length. These variables have shown to change according to level of experience. Mackay, Datta & Chang et al[59] advocate a skills examination as a way of assessing technical ability which has been applied in both the UK and Hong Kong.[60] Their multiple objective measures of skills included six tasks: knowledge of sutures and instruments, knowledge of surgical devices, knot formation, skin-pad suturing, excision of a skin lesion and laparoscopic manipulation.

The specialties are very diverse and different skills would be required for different specialties. The surgical skills mentioned above would appear to be of little importance in, for example, a specialty such as psychiatry. Perhaps the ability to engage emotionally may be a relevant skill to test for those entering psychiatry and a low tendency for confirmation bias for those entering general practice. However, there has been little research that would enable recommending such tests in selection.

Work sample

Work sample tests are simulations of part, or all, of the job that must be performed by applicants. They are a set of actual tasks that are physically and/or psychologically similar to those in the job. These types of tests are most often used to hire skilled workers such as welders, machinists and carpenters. It may involve, for example, repairing a series of defective motors. This type of test would seem generally applicable to the medical profession due to the level of training and job knowledge one must have. They are viewed by managers to be the most valid predictor of performance, and studies have concluded that they have lower levels of standardised ethnic group differences and adverse impact, and are viewed in a positive light by applicants.[61]

Older studies and reviews have reported work sample tests as having the highest predictive validity of all selection methods.[11] This is not unsurprising as prior performance is a predictor of future performance. However, a more recent meta-analysis disputes the high predictive validity of work sample tests.[61] This meta-analysis updates previous work and excludes studies that are considered to be methodologically flawed (such as, including work sample tests that are more appropriately defined as job knowledge tests or a situational interview), and concluded that work sample tests have a lower predictive validity than previously given.

When deciding what to include in any selection process a number of factors need to be considered and trade-offs often need to be made. While work sample tests may have lower predictive validity than originally claimed, there are other positive attributes that may still warrant their usage, particularly their claimed lower levels of bias with regard to ethnicity. For selection to specialty training, work sample tests may be of limited use. Individuals would still be in training and would not have been exposed to all the tasks and skills of a particular specialty. It would appear that work sample tests in this context would be counterproductive; the test would assess something that the individual has not yet learned, practised and mastered.

Combining methods and assessment centres

As discussed above, general mental ability (GMA) provides the greatest predictive value over other selection measures. In fact, the power of GMA tests increases as job complexity increases.[62, 63] The predictive validity of GMA may also be enhanced by combining it with other selection tools. Table 1 and table 2 below illustrate how GMA tests can have greater predictive value when used in conjunction with other measures. The combinations with the highest validity for job performance are GMA and a work sample test, GMA plus an integrity test and GMA plus a structured interview.

As discussed previously the inclusion of measures of GMA in selection for specialty training, or GMA as part of an assessment centre for that matter, may be redundant if it is also tested for at admission to medical school. Assessment centres are commonly used in the recruitment of managerial level positions and graduates internationally. These centres usually combine a number of the tools discussed above and group activities such as leaderless group discussion or problem solving exercises. Assessment centres may have substantial validity but only moderate incremental validity over GMA. This is in part due to the case that most assessment centres include a measure of GMA. The study of assessment centres is somewhat problematic in that it is not often clear what is measured by assessment centres.[11, 64]

Table 1. Predictive validity for overall job performance of GMA scores combined with a second predictor using (standardised) multiple regression (adapted)[11]

Selection tools	Validity (r)	Multiple R	Gain in validity from adding supplement	% increase in validity	Standardised regression weights	
					GMA	Supplement
GMA tests	.58					
Work sample tests	.54	.63	.12	24%	.36	.41
Integrity tests	.41	.65	.14	27%	.51	.41
Conscientious tests	.31	.60	.09	18%	.51	.31
Interviews (structured)	.51	.63	.12	24%	.39	.39
Interviews (unstructured)	.38	.55	.04	8%	.43	.22
Job knowledge tests	.48	.58	.07	14%	.36	.31
Reference checks	.26	.57	.06	12%	.51	.26
Job experience	.18	.54	.03	6%	.51	.18
Biographical data	.35	.52	.01	2%	.45	.13
Assessment centres	.37	.53	.02	4%	.43	.15
Years of education	.10	.52	.01	2%	.51	.10
Interests	.10	.52	.01	2%	.51	.10
Graphology	.02	.51	.00	0%	.51	.02
Age	-.01	.51	.00	0%	.51	-.01

Table 2. Predictive validity for overall performance in job training programmes of selection tools[11]

Selection tools	Validity (r)	Multiple R	Gain in validity from adding supplement	% increase in validity	Standardised regression weights	
					GMA	Supplement
GMA tests	.56					
Integrity tests	.38	.67	.11	20%	.56	.38
Conscientious tests	.30	.65	.09	16%	.56	.30
Interviews	.35	.59	.03	5%	.59	.19
Peer ratings	.35	.57	.01	1.4%	.51	.11
Reference checks	.23	.61	.05	9%	.56	.23
Job experience	.01	.56	.00	0%	.56	.01
Biographical data	.30	.56	.00	0%	.55	.03
Years of education	.20	.60	.04	7%	.56	.20
Interests	.18	.59	.03	5%	.56	.18

Patterson, Ferguson & Norfold et al[52] are investigating, in a longitudinal study, the predictive validity of an assessment centre for general practice registrars. This assessment centre includes a simulated consultation, a group exercise where candidates are asked to resolve a work related issue, an in-tray exercise requiring the justified prioritisation of work, a competency-based structured interview and a medical interview. The entire selection procedure also includes an application form of biographical information, six structured competency questions and personal statements and referees ratings of candidates on six competencies. Preliminary findings, after three months, indicate that the assessment centre had greater predictive validity over traditional methods. This holds interesting possibilities for general practice and other specialties.

The combination of different tools for selection is attractive for a number of reasons. It allows the triangulation of information and the assessment of different knowledge, skills, ability and attitudes of an individual in different ways and situations. The importance of the roles the trainees will take up and the diverse attributes that are required suggest that a multimethod approach would be ideal.

Is testing valid and reliable?

The following is discussed with reference to personality testing, but applies generally to other tests, for example, aptitude and ability tests.

Schmidt & Hunter[11] in their meta-analysis of 85 years of personnel psychology research concluded that general mental ability (GMA) tests have the highest validity and lowest application cost. The research evidence for the validity of GMA measures for predicting job performance is stronger than most other methods (see table 3), and is also the best predictor of acquisition of job knowledge on the job (training performance, see table 4). These results are supported by more recent research in the UK, which specifically incorporates health professionals, including surgeons, as one of the occupational categories investigated.[63]

Table 3. Predictive validity for overall job performance of selection tools (adapted)[11]

Selection tool	Validity (r)
GMA tests	.51
Work sample tests	.54
Integrity tests	.41
Conscientious tests	.31
Interviews (structured)	.51
Interviews (unstructured)	.38
Job knowledge tests	.48
Reference checks	.26
Job experience	.18
Biographical data	.35
Assessment centres	.37
Interests	.10

Table 4. Predictive validity for overall performance in job training programmes of selection tools (adapted)[11]

Selection tool	***Validity (r)***
GMA tests	*.56*
Integrity tests	*.38*
Conscientious tests	*.30*
Interviews	*.35*
Peer ratings	*.35*
Reference checks	*.23*
Job experience	*.01*
Biographical data	*.30*
Interests	*.18*

Validity

It may be unclear what variables should be measured in a personality test. Validity[l] is simply whether the test measures what it wants to measure. Empirical approaches have been used in the development of personality tests to determine those factors one would wish to measure. These have resulted in what is sometimes referred to as the big five personality factors of emotional stability, extroversion, agreeableness, conscientiousness, and openness to experience. There are other competing models which vary from four to 16 factors. Another way of test construction builds on a personality theory and testing the variables within it, which leads to different results again.

It can be difficult to pool together and write the items that adequately represent the factors being measured. A personality trait is complex and trying to capture all the subtleties within it in a number of questions is challenging. Settling on item response also needs to be considered. Often dichotomous Yes-No, Like-Dislike or True-False responses are deemed too extreme and may frustrate respondents who cannot give a definitive response. On the other hand, including a middle response can increase the likelihood of respondents choosing this and rendering data ineffectual. Some individuals also have a tendency to respond in more extreme or uncertain middle-ground ways which may make measuring behaviour difficult. Further, quantifiers used in items may be too vague eliciting differing responses from different individuals and rendering test scores useless. For example, what does ‘few’ or ‘often’ mean in the minds of different respondents? Items should avoid terms of frequency, refer to specific rather than general behaviour and be clear and unambiguous.[35] Individuals not responding honestly, or answering questions as they think they should be answered, are also criticisms of personality tests. Some tests incorporate 'lie scales' which are designed to identify individuals responding in socially desirable ways. Personality tests also rely on self-reports of behaviour, which are not the same as the behaviour to which they refer. As stated previously, individuals may lie, may be mistaken and may interpret terms differently from the test constructor and from other individuals.

---

[l] Face validity – appears to be measuring what it claims to measure.
Concurrent validity – a test’s correlations with other similar tests taken at the same time. Where other benchmark tests exist correlations above .75 are considered to be good support of concurrent validity, and the test must have some advantage compared with the test it is being compared to .35
Predictive validity – it predicts some criterion or criteria.
Content validity – if items judged by experts in the field are suitable for their purpose (usually for an achievement or ability test).
Incremental and differential validity – ability of a test to predict one criterion better than another test.
Construct validity – if the results of the test fit hypotheses concerning the nature of the test variable.

Reliability

Reliability[m] refers to the stability of test scores over time and the internal consistency of a test. A number of factors influence a test's reliability. These include:

- changes in the individual (for example, when testing transient moods such as anger, reliability between tests might be low)
- impaired performance (for example, due to sickness, domestic or emotional problems or fatigue)
- test-taking environment (for example, uncomfortable seat, accidentally missing out test pages or pen running out of ink)
- poor test instructions
- subjective scoring (those scoring the test arrive at different scores)
- guessing
- time gap (a three-month lapse between taking the same test is recommended)[35]
- sample size (the larger the sample the smaller the statistical error. Samples should contain at least 100 participants).

It must be emphasised that test results should not be seen in isolation and assessments of students for selection should be based on information from a variety of sources.

---

[m] Test-retest reliability – the correlation of scores from a set of participants who take the test on two occasions. Test-retest reliability should be high .35
Internal consistency reliability – how well the items that measure the same construct yield similar results.

Conclusion

Selecting trainees for specialty training is a complex task. A multitude of factors need to be considered including criteria for selection (potential, diversity and for what outcomes), the context in which selection fits (that is, other selection points and assessments) and the effectiveness and efficiency of selection methods available. It is particularly pertinent to address these now because a redesigned selection process will need to be in place for the first cohort of foundation programme trained doctors entering specialty training in 2007. Combining selection methods appears to be the most favourable approach because no one selection method is supported by enough research to be solely relied on in the medical field. Further, combining certain selection methods can result in incremental predictive validity.

In asking important questions such as ‘is there an optimal doctor personality?’, ‘how do we test for potential?’ and ’are tests reliable and valid?’, it is clear that there is no simple answer. What is clear is that more research is needed in all areas of selection design from competency modelling to predictive validity of tests specifically for UK doctor samples. While it is encouraging that work has started in some specialties and at an undergraduate level, there appears to be a long way to go before some questions can be answered. Trainability and performance in job training programmes as outcomes has been studied in other industries and medical education can learn from this research.

The cost of getting selection for specialty training wrong is high. As well as financial, these costs are both public – in terms of the service a doctor provides throughout their medical career, and personal – to the doctor who spends a considerable amount of time and effort working towards the ultimate goal of independent practice as a consultant or general practitioner. Getting selection right, therefore, is in everyone’s interest. Implementation of selection systems should be carefully researched and considered to help ensure the objective of competent, well-trained and suitable doctors is met.

References

1 PMETB (2006) *PMETB principles for entry to specialist training*. www.pmetb.org.uk/media/pdf/5/m/PMETB_Press_release_-_31_January_2006_-_PMETB_publishes_principles_for_entry_to_specialist_training.pdf (accessed November 2006)

2 UK departments of health (1998) *A guide to specialist registrar training.* Department of Health.

3 www.mmc.nhs.uk (accessed November 2006)

4 MMC & BMA (2005) *The rough guide to the foundation programme.* London: The Stationery Office.

5 GMC (2003) *Tomorrow's doctors*. London: General Medical Council.

6 Richards P, Stockill S, Foster R et al (2006) *Learning medicine* (17e). Cambridge: Cambridge University Press.

7 Schwartz S (2004) *Admissions to Higher Education Review. Fair admissions to higher education: Recommendations for good practice.* London: Admissions to Higher Education Steering Group.

8 McManus IC , Powis DA & Wakeford R et al (2005) Intellectual aptitude tests and A levels for selecting UK school leaver entrants for medical school. *British Medical Journal* **331**: 555-9.

9 www.ukcat.ac.uk (accessed November 2006)

10 www.bmat.org.uk (accessed November 2006)

11 Schmidt FL & Hunter JE (1998) The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin* **124**: 262-74.

12 Posthuma RA, Morgeson FP & Campion MA (2002) Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology* **55**: 1-81.

13 Ferguson E, James D & O'Hehir F et al (2003) Pilot study of the roles of personality, references and personal statements in relation to performance over five years of a medical degree. *British Medical Journal* **326**: 429-32.

14 Henley S & Stevenson H (1989) *The role of the surgeon: Report on a job analysis commissioned by the Royal College of Surgeons.* Thames Ditton: Saville & Holdsworth (UK).

15 www.dh.gov.uk/PolicyAndGuidance/HumanResourcesAndTraining/ModernisingPay/AgendaForChange/fs/en (accessed November 2006)

16 Gilligan JH, Treasure T & Watts C (1996) Incorporating psychometric measures in selecting and developing surgeons. *Journal of Management in Medicine* **10**: 5-16.

17 Barnes RW (1987) Surgical handicraft: Teaching and learning surgical skills. *The American Journal of Surgery* **153**: 422-7.

18 McLachlan JC (2005) Outreach is better than selection for increasing diversity. *Medical Education* **39**: 872-5.

19 GMC (2005) *The new doctor.* London: General Medical Council.

20 GMC (2006) *Good medical practice*. London: General Medical Council.

21 Firth-Cozens J (1999) What makes a good doctor? Assessing personal attributes. In, British Medical Association, *Selecting our doctors: A report on a BMA conference on selection for medical school.* London: British Medical Association.

22 Pritchard L (2005) Can a test weed out rogue doctors? *Student BMA News*, February p 4.

23 Bore M, Munro D, Kerridge I et al (2005) Selection of medical students according to their moral orientation. *Medical Education* **39**: 266-75.

24 Firth-Cozens J, Cording H & Ginsburg R (2003) Can we select health professionals who provide safer care. *Quality & Safety in Health Care* **12**: 16-20.

25 Kirk LM & Blank LL (2005) Professional behaviour: A learner's permit for licensure. *New England Journal of Medicine* **353**: 2709-11.

26 Papadakis MA, Teherani A & Banach MA et al (2005) Disciplinary action by medical boards and prior behavior in medical school. *New England Journal of Medicine* **353**: 2673-82.

27 Grant J, Chambers K & Maxted M et al (2005) *Modernising medical careers: Identification of career intentions as a function of foundation programme experience.* Milton Keynes: Open University Centre for Education in Medicine.

28 McParland M, Noble L & Livingston G et al (2003) The effect of a psychiatric attachment on students' attitudes to and intention to pursue psychiatry as a career. *Medical Education* **37**: 447-54.

29 Gale R & Grant J (2002) Sci45: The development of a specialty choice inventory. *Medical Education* **36**: 659-66.

30 Petrides KV & McManus IC (2004) Mapping medical careers: Questionnaire assessment of career preferences in medical school applicants and final-year students. *BMC Medical Education* **4**: 18-34.

31 Stilwell NA, Wallick MM & Thal SE et al (2000) Myers-Briggs type and medical specialty choice: A new look at an old question. *Teaching and Learning in Medicine* **12**: 14-20.

32 Department for Education and Skills (2003) *The future of higher education*. London: The Stationery Office.

33 UCAS (2005) *UCAS data.* www.ucas.com/figures/data/data.html (retrieved 24 March 2006)

34 BMA (2004) *The demography of medical schools: A discussion paper.* London: British Medical Association.

35 Kline P (2000) *Handbook of psychological testing* (2e). London: Routledge.

36 Brown CA, Wakefield SE & Bullock AD (2003) The selection of GP trainees in the West Midlands: second audit of assessment centre scores by ethnicity and country of qualification. *Medical Teacher* **25**: 649-53.

37 Stokes GS (1999) Introduction to special issue: The next one hundred years of biodata. *Human Resource Management Review* **9**: 111-16.

38 McManus IC (1999) Non academic factors in selection. In, British Medical Association, *Selecting our doctors: A report on a BMA conference on selection for medical school.* London: British Medical Association.

39 Pritchard L (2006) Are you playing the game? *Student BMA News*, January p 4.

40 Taylor I (2005) Can performance as an undergraduate assist entry selection into surgical training programmes? [Editorial]. *Annals of the Royal College of Surgeons of England* **87**: 1-2.

41 Graves L & Karren RJ (1996) The employee selection interview: A fresh look at an old problem. *Human Resource Management* **35**: 163-80.

42 Chapman D S & Zweig DI (2005) Developing a nomological network for interview structure: Antecedents and consequences of the structured selection interview. *Personnel Psychology* **58**: 673-702.

43 Campion MA, Palmer DK & Campion JE (1997) A review of structure in the selection interview. *Personnel Psychology* **50**: 655-702.

44 Smith M & George D (1994) Selection methods. In, Cooper CL & Robertson IT (eds), *Key Reviews in Managerial Psychology*. Chichester: John Wiley.

45 Maurer SD (1997) The potential of the situational interview: Existing research and unresolved issues. *Human Resource Management Review* **7**: 185-201.

46 Rudman LA, Greenwald AG & Mellott DS et al (1999) Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition* **17**: 437-65.

47 van Leeuwen YD, Mol SSL & Pollemans MC et al (1997) Selection for postgraduate training for general practice: The role of knowledge tests. *British Journal of General Practice* **47**: 359-62.

48 Costa PT & McCrae RR (2006) Age changes in personality and their origins: Comment on Roberts, Walton and Viechtbauer (2006). *Psychological Bulletin* **132**: 26-8.

49 Roberts BW, Walton KE & Viechtbauer W (2006) Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin* **132**: 1-25.

50 Gilligan JH, Welsh FKS & Watts C et al (1990) Square pegs in round holes: Has psychometric testing a place in choosing a surgical career? A preliminary report of work in progress. *Annals of the Royal College of Surgeons of England* **81**: 73-79.

51 Fitzgerald C (1993) *Royal College of Surgeons validation study.* Thames Ditton: Saville & Holdsworth (UK).

52 Patterson F, Ferguson E & Norfold T et al (2005) A new selection system to recruit general practice registrars: Preliminary findings from a validation study. *British Medical Journal* **330**: 711-4.

53 Scott PV (2000) Politically correct references [Personal view]. *British Medical Journal* **320**: 1613.

54 Taylor PJ, Pajo K & Cheung GW et al (2004) Dimensionality and validity of a structured telephone reference check procedure. *Personnel Psychology* **57***:* 745-72.

55 Bulstrode C & Hunt V (2003) Selecting the best from the rest. *The Surgeon - Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland* **1**: 328-31.

56 Bann S & Darzi A (2005) Selection of individuals for training in surgery. *American Journal of Surgery* **190**: 98-102.

57 Francis NK, Hanna GB & Cresswell AB et al (2001) The performance of master surgeons on standard aptitude testing. *American Journal of Surgery* **182**: 30-3.

58 Bann SD, Khan MS & Darzi AWD (2003) Measurement of surgical dexterity using motion analysis of simple bench tasks. *World Journal of Surgery* **27**: 390-4.

59 Mackay S, Datta V & Chang A et al (2003) Multiple objective measures of skill (MOMS): A new approach to the assessment of technical ability in surgical trainees. *Annals of Surgery* **238**: 291-300.

60 Bann S, Kwok KF & Lo CY et al (2003) Objective assessment of technical skills of surgical trainees in Hong Kong. *British Journal of Surgery* **90**: 1294-9.

61 Roth PL, Bobko P & McFarland LA (2005) A meta-analysis of work sample test. *Personnel Psychology* **58**: 1009-37.

62 Hunter JE & Hunter RF (1984) Validity and utility of alternative predictors of job performance. *Psychological Bulletin* **96**: 72-98.

63 Bertua C, Anderson N & Salgado JF (2005) The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology* **78**: 389-409.

64 Garavan TN & Morley M (1998) Graduate assessment centres: An empirical investigation of effectiveness. *Education and Training* **40**: 206-19.

Department of Science and Education
British Medical Association, BMA House, Tavistock Square, London, WC1H 9JP
www.bma.org.uk
E info.science@bma.org.uk

© British Medical Association, 2006

CODE: BMA 47187 / AB / BOARD OF MEDICAL EDUCATION / 11.06