Journal of Educational Evaluation for Health Professions

J Educ Eval Health Prof 2013, 10: 2 • http://dx.doi.org/10.3352/jeehp.2013.10.2

Open Access

REVIEW

Assessment methods in surgical training in the United Kingdom

Evgenios Evgeniou[1]*, Loizou Peter[2], Maria Tsironi[3], Srinivasan Iyer[1]

[1]Department of Plastic Surgery, Wexham Park Hospital, Slough, United Kingdom; [2]Department of Ear, Nose and Throat, Luton & Dunstable Hospital, Luton, United Kingdom; [3]Department of Ear, Nose and Throat, Wrexham Maelor Hospital, Wrexham, United Kingdom

Abstract

A career in surgery in the United Kingdom demands a commitment to a long journey of assessment. The assessment methods used must ensure that the appropriate candidates are selected into a programme of study or a job and must guarantee public safety by regulating the progression of surgical trainees and the certification of trained surgeons. This review attempts to analyse the psychometric properties of various assessment methods used in the selection of candidates to medical school, job selection, progression in training, and certification. Validity is an indicator of how well an assessment measures what it is designed to measure. Reliability informs us whether a test is consistent in its outcome by measuring the reproducibility and discriminating ability of the test. In the long journey of assessment in surgical training, the same assessment formats are frequently being used for selection into a programme of study, job selection, progression, and certification. Although similar assessment methods are being used for different purposes in surgical training, the psychometric properties of these assessment methods have not been examined separately for each purpose. Because of the significance of these assessments for trainees and patients, their reliability and validity should be examined thoroughly in every context where the assessment method is being used.

Key Words: *Assessment; Examination; Recruitment; Surgical training*

INTRODUCTION

Medicine is a satisfying and financially rewarding profession that demands multiple physical and cognitive skills as well as a stable personality with appropriate traits. The aspiring surgeon, from the point of application to a programme of medical studies, is committed to a long journey of assessment that evaluates the acquisition of the knowledge, skills, and attributes expected of a developing or practicing doctor. The various assessment methods have significant implications for trainees, such as not permitting them to enter or progress in their chosen specialty, as well as on the general public, when allowing non-competent doctors to progress or practice following certification. Therefore, appropriate assessment methods, with proven reliability, validity, and feasibility, must be established for selection of candidates to medical school, job selection, progression in training, and certification. This review attempts to investigate the various assessment methods used for the above purposes in surgical training and examine the available evidence in the literature regarding their psychometric properties.

THE PSYCHOMETRIC PROPERTIES OF ASSESSMENT METHODS

The psychometric properties of an assessment method are the characteristics that describe how well an assessment method can evaluate what it is designed to evaluate, typically including its validity and reliability. Validity describes how well an assessment measures what it is designed to measure, and it is subdivided into different types. Face validity refers to the functionality and realism of a test. Content validity refers to whether a test is suitable as a measure of what it is designed to

*Corresponding email: evgenios@doctors.org.uk

Received: November 25, 2012; Accepted: January 31, 2013;
Published: February 5, 2013
This article is available from: http://jeehp.org/

measure, and construct validity is an indicator of whether an assessment is successful in measuring what it is supposed to measure. Incremental or criterion validity is a comparison of tests that measure the same trait. Predictive validity or outcome validity is the ability of a test to predict future performance in a specific domain [1,2].

Reliability informs us whether a test is consistent in its outcome by measuring the reproducibility and discriminating ability of a test. In order to assess the reliability of a test we use various items, such as inter-test reliability, which shows if the assessment gives the same result if repeated, inter-rater reliability, which refers to the agreement of scores given by different raters on the same subject, and internal consistency, which reflects the correlation of the different items of a test and their contribution to the outcome of the test [1,2]. Reliability ranges from 0 to 1, with a measurement of 0.8 being appropriate for high-stakes assessment [2].

Assessment methods can be summative or formative. Formative assessments are of an informative nature, are used to provide feedback, and aim at development, while summative assessments are used for selection [1]. The results of a summative assessment can be based on norm-referencing or criterion referencing [2,3]. In norm-referencing, each result is compared with the other results from the same cohort, and the ranking of the test participants is used to distribute grades and make decisions regarding selection or pass/fail (for example, only the top ranked 30% of the test participants may be selected to pass). During criterion-referencing, the result is judged against a predefined standard or a set of criteria. For example, the exam candidate must demonstrate the minimum ability in a certain domain, such as clinical examination, in order to be judged as fit to practice and therefore pass an exam. However, the performance of the other exam participants is not taken into account.

A SURGEON'S JOURNEY OF ASSESSMENT

Selection into a program of study

The study of medicine requires a threshold academic ability and a well rounded personality with characteristics suitable for a career in medicine, such as motivation, integrity, communication skills, empathy, decision-making ability, teamwork, and self-awareness [4,5]. The selection process for the study of medicine in the United Kingdom comprises a combination of assessment methods which test the applicants' cognitive and non-cognitive traits. The cognitive criteria have traditionally been assessed by previous academic performance in the form of General Certificate of Secondary Education (GCSE) scores and predicted A-level scores. Intellectual amplitude tests, such as the Biomedical Admissions Test (BMAT) and the UK Clinical Aptitude Test (UKCAT), are tests which measure performance across a range of mental abilities and are used to predict future performance in education programmes. These tests have been introduced in an attempt to make the selection process fairer, increase the diversity of students, and assist in the selection process from a growing number of applicants with a similar level of academic achievement [6]. The non-cognitive criteria are assessed by the Universities and Colleges Admissions Service (UCAS) application form, including the applicant's personal statement and reference letter, and by an interview [5].

Job selection

Surgical training in the UK has shifted from a traditional apprenticeship model to a competency-based model. According to this model of training, trainees cannot progress and cannot complete their training if they cannot demonstrate competence in predefined areas of the curriculum. For example, a surgical trainee must prove his/her ability to examine a patient, form a deferential diagnosis, and manage a case of appropriate complexity in order to be able to progress to the next level of training. Changes are constantly being made in the recruitment process for postgraduate surgical training in order to make it fair, eliminate discrimination, and choose applicants that are competent for the job [7]. Before the selection of any assessment method for the process of recruitment, it is important to perform a job analysis in order to identify the competencies required for a certain specialty. According to Patterson et al. [8], we need to take into account not only clinical knowledge and academic achievement, but also a wide range of attributes, both common to all specialties and specialty specific, during the selection process. The selection process for postgraduate surgical training in the UK starts with an application form which includes the curriculum vitae (CV), elements of past achievements and clinical experience, demographic information, and focused questions. The interviews, which are usually structured, assess mainly the candidate's CV, referee reports, and portfolio. Assessments centres that combine interview, portfolio assessments, and work-related task stations have been successfully used in the recruitment process in surgery, as well as in as in other specialties such as paediatrics and anaesthesia [7,9].

Progression

Postgraduate training in the UK starts with the foundation programme (Fig. 1). The goals of the foundation programme are to determine fitness to progress to the next level of training, provide focused feedback to trainees for their development, identify doctors who may face difficulties in their everyday practice or training, and have assessment methods with

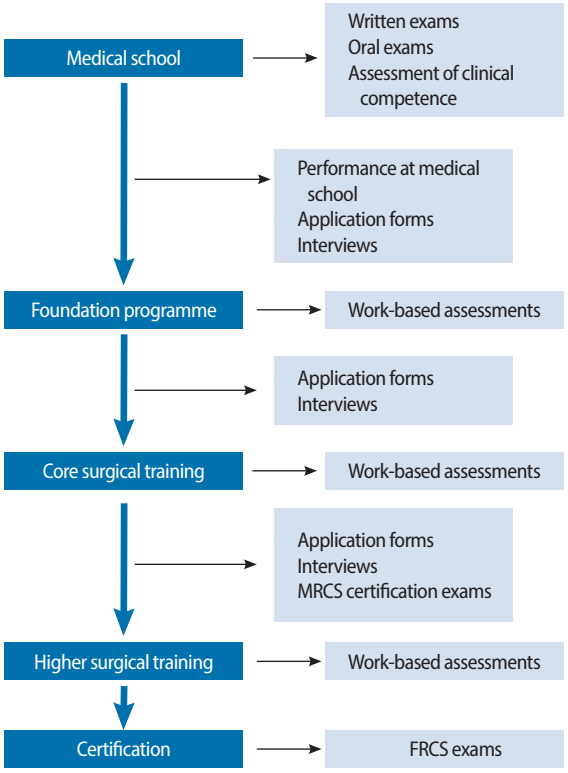

Fig. 1. The surgical training pathway. MRCS, Member of the Royal College of Surgeons; FRCS, Fellow of the Royal College of Surgeons.

the appropriate psychometric properties to guarantee patient safety [10]. Following the satisfactory completion of the foundation programme and acquisition of the foundation competencies, the aspiring surgeon enters surgical training starting with core surgical training. At this stage, the trainee acquires the basic principles of surgery in general, and continues with training in his/her chosen surgical subspecialty (such as general surgery or orthopaedics).

Work based assessments (WBAs) have been introduced in foundation and specialty training to assess the "does" level of Miller's pyramid [11,12]. The main aims of WBAs are to aid learning through objective feedback and to assess curriculum competencies [11]. Some of the assessments (mini-clinical evaluation exercise (mini-CEX), case-based discussion (CBD), mini peer assessment tool [mini-PAT]) are common to foundation and specialty training, while others (surgical direct observation of procedural skills [S-DOPS] and procedure-based assessments [PBA]) are specific to surgical specialty training [13,14]. The mini-CEX is a record of trainee-patient interaction observed by an assessor. The CBD is an evaluation of the trainee's performance during the clinical case and is usually based on a review of patient case notes. The S-DOPS is a record of direct observation of a practical skill performed by the trainee which is usually aimed at junior trainees. PBA are records of direct observation of more complex procedures performed in the operating theatre, which are more appropriate for senior trainees. The mini-PAT uses multi-source feedback from a variety of healthcare professionals to assess the trainees' professional and behavioural skills, such as communication skills, team-work, judgment, compassion, and probity. Surgical logbooks of operations have been used in specialty training as an indicator of acquired experience and engagement with training [2,11]. The trainees performance during a training post is evaluated by a committee assigned by the Deanery in an annual review of competence progression (ARCP), which uses a variety of assessment methods, such as WBAs, logbooks, and supervisor reports in order to assess the trainees competence to progress to the next level of training [13]. Although these assessments were designed for formative purposes to provide feedback on the performance of trainees, they are currently being used for summative purposes, both as criteria for progression in training and in job selection. Both in the foundation programme and during specialty training, trainees are expected to keep a portfolio. Portfolios should mirror the trainee's achievements, and they have been used in postgraduate training to provide summative assessment and encourage reflective practice [15,16].

Certification

Specialty certification in the UK is regulated by the Royal Colleges and takes the form of multi-stage exams. These exams are usually criterion-referenced, requiring a baseline level of competency in order to grant certification [17]. Surgical certification exams have been established in order to safeguard patients and ensure high standards for practising surgeons [18]. The Member of the Royal College of Surgeons (MRCS) exam takes the form of a summative assessment which assesses the acquisition of the knowledge, skills, and attributes required for completion of the core training. This allows progression to higher specialist training [17]. Part A of the exams, which usually comprises of multiple choice/extended matching questions (MCQ/EMQ), tests whether the candidate has adequate basic science knowledge before testing the application of knowledge in a clinical context using vivas and objective structured clinical examination (OSCE) methods.

The purpose of the Fellow of the Royal College of Surgeons (FRCS) exam is to assess whether the candidate has achieved a desirable level of knowledge and skills following the completion of his training, thereby certifying that the candidate has achieved the standards of a trained surgeon and is ready to practice safely as a consultant [19]. The senior surgical trainee approaching the completion of his training must be able to demonstrate sufficient knowledge, judgement, and experience

in order to be allowed to practise independently [18]. Similar to the MRCS, the FRCS comprises a written exam, the successful completion of which allows progression to the next stage of the examination. The second stage of the exam uses long cases and vivas to assess the clinical competence of the candidate [19,20].

SIMILAR ASSESSMENTS FOR DIFFERENT PURPOSES

Written tests

Written tests are a very common assessment method and are used for selection and certification purposes. Written tests, such as MCQs and short-answer questions (SAQs), although designed to test factual knowledge, can also be used to test the application of knowledge if they are carefully designed. MCQs have high reliability because of the large number of testing items and the standardised way of marking [2,21] and are therefore very popular in high stakes examinations.

Written tests combining short essay questions, MCQs, EMQs, and rating questions have been shown to be successful shortlisting tools for Core Medical Training and General Practice training selection processes. These tests have shown high reliability, high predictive validity for subsequent interview and selection centre scores, high incremental validity, and cost effectiveness compared with other shortlisting methods [22,23]. The use of standardised marking techniques, such as machine marking, increased validity, and efficiency, has revealed the shortcomings of short essay type questions, however.

Application forms are used for shortlisting purposes for selection into a programme of study or job selection. They frequently use short essay type questions, mainly in the form of statements as prompts. Although these types of assessments have shown predictive validity regarding performance in medical school and at selection centres [5,23], they are unreliable because of uncontrolled variance in the time needed for completion and external influence, such as the internet, and are very difficult to mark [24,25].

Although written amplitude tests are being used in selection to medical schools worldwide, there are conflicting opinions in the literature regarding their psychometric properties. The Medical College Admissions Test (MCAT), which is used in the United States, has shown predictive validity for performance in licencing examinations and medical school grades. Regarding the two main amplitude tests used by universities in the UK, some studies have demonstrated good reliability and predictive validity for year 1 and 2 medical school examinations for UKCAT and predictive validity for the pre-clinical years performance for BMAT [25-27]. However, other authors have questioned the reliability and incremental validity of the BMAT compared with other measures of scientific knowledge such as GCSEs and A-levels and have demonstrated that the UKCAT does not predict performance in year 1 of medical school [26,28].

Although there is very little evidence regarding the validity of the use of written assessment methods for specialty certification purposes, MCQ and EMQ tests are very popular assessment methods for postgraduate examination purposes because of their high reliability and feasibility when utilised for other purposes, as demonstrated above. The psychometric properties, though, are different when the assessment methods are being used for the purpose of selection compared to when used for the purpose of certification, and their reliability and validity must be demonstrated and not assumed for high stakes certification exams.

Vivas and orals

Oral examinations are frequently used for certification purposes, such as the MRCS and FRCS specialty examinations. Vivas have been criticised in the literature for having low reliability and validity for high stakes examinations and a very high cost [19,29]. The low reliability of oral examinations is attributed to the introduction of personal bias through active participation of the examiner in the exam. Also, although oral examinations have the advantage of flexibility of moving from one topic to another, the lack of standardisation due to the varying content, the level of difficulty, and the level of prompting for each candidate, reduces reliability [29]. A candidate's appearance, verbal style, and gender have been shown to influence oral examination scores, creating concerns regarding discrimination. Davis and Karunathilake [29], in a review of the literature on oral examinations, concluded that one of the disadvantages is that testing is usually done at a low taxonomic level according to Bloom' s cognitive domain taxonomy, which provides a sequential classification of levels of thinking skills (Fig. 2). Assessment tends to remain at the level of factual knowledge, without testing higher order problem-solving and decision-making. Other authors [19], though, have noted that this is mainly due to the examiners not utilising the potential of vivas to demand higher order thinking. Indeed, one of the claimed advantages of oral examinations is that they offer the opportunity of questioning in depth, although this advantage is underused because of the time restrictions in oral examinations and the generally low taxonomic level of questioning [29].

Various suggestions have been made in order to reduce bias and increase the reliability of oral examinations, such as training the examiners, using multiple orals and multiple examiners, standardisation of questions and using descriptors and criteria for marking the answers [19,29]. These measures, however, would increase the resource requirements and costs for

Fig. 2. Bloom's cognitive domain taxonomy (adapted from: http://ww2.odu.edu/educ/roverbau/Bloom/blooms_taxonomy.htm).

oral examinations, creating concerns regarding feasibility, especially when other assessment methods are available to test the same domains. Iqbal et al. [19], concede that oral examinations are costly and resource-intensive, but emphasize that they are unique in providing a global impression of the candidates where personality, professionalism, and operational knowledge can be better assessed than through other methods.

Interviews and portfolios

Interviews are used for selection into a programme of study and job selection. However, because of the active participation of the interviewer in the assessment process and the introduction of personal biases, interviews have similar reliability concerns as with oral examinations [5]. Measures to increase reliability are similar to oral examinations and include training the interviewers, using multiple stations and multiple interviewers, and standardisation of the interview questions and scoring methods [30]. Studies have shown that interviews designed by taking into account these factors, such as the multiple mini interviews (MMI) used in the undergraduate and postgraduate selection processes, can achieve very high reliability [31,32]. Interviews have not been shown to have adequate predictive validity for academic achievement, however [5,28,33]. On the other hand, assessment centres in the interview selection process for postgraduate training selection, which are interview stations that assess the specific skills and competences previously identified by a thorough job analysis, have shown to have high predictive validity for future job performance [7,9,34].

Portfolios, as a record of achievements and experiences, are being used in the selection of candidates for post-graduate training, assessment of fitness to progress in training, and in revalidation. The use of portfolios for summative assessment purposes has been criticised as lacking in reliability and validity because of the difficulty in extrapolating quantitative data from portfolios [16,35]. Some authors have suggested triangulating portfolio data with other assessment methods, using global criteria with standards of performance (rubrics), training the assessors, and using multiple raters and discussion between raters in order to improve the reliability of the use of portfolios for summative purposes [15,16]. For the effective use of portfolios, mainly for formative purposes, there should be clear guidelines for both trainees and assessors and specific portfolio goals, but caution has to be taken not to become too descriptive, so that they do not lose their reflective and creative character [15,16].

Assessment of clinical competence

Assessment methods based on the direct observation of clinical and procedural skills are being used for formative purposes in postgraduate and undergraduate training and for summative purposes in certification exams. Various tools have been developed to assess the different aspects of clinical practice, and different tools are being used to assess competence to progress in training compared to assessment of clinical competence for certification.

Research has shown that the mini-CEX, CBD, DOPS, and multi-source feedback, in the form of mini-PAT, are feasible, reliable, and valid assessment methods, with their results correlating with other assessment formats (criterion validity) and able to differentiate between different levels of competence (construct validity) [12,14,36]. Direct observation of procedural skills using the PBA form during real procedures, and the objective structured assessment of technical skills (OSATS) form during simulated procedures, have also been shown to have good reliability and validity [2].

Long cases have been used for the assessment of clinical skills in both undergraduate and postgraduate training, but also for certification purposes. Long cases can be either observed or unobserved and based on the candidate's presentation of the case. In order to achieve a reliability level appropriate for high stakes examinations, it has been suggested that long cases should be observed and have adequate length, or alternatively multiple shorter cases should be used [3].

OSCEs have been used for the assessment of clinical competence for certification purposes, such as medical school finals and MRCS, and have recently been used for job selection purposes in assessment centres. The high validity and reliability of OSCE examinations, which makes it an appropriate assessment format for high stakes examinations, is based on objectivity, standardisation, and authenticity in recreating real clinical circumstances [3,37]. Measures to increase the reliability of OSCEs include careful sampling across different do-

Table 1. Assessment methods in surgical training

Assessment method		Description	Applications
Written tests			
Other		MCQs, EMQS, and SAQs are well established written exam for-mats used for multiple purposes	Selection to a programme of training certification
GCSEs		Exams on selected subjects depending on the chosen field of future higher education	Selection to medical school
Application form		Depending on the purpose they combine different elements, such as a CV and short essay type questions.	Selection to a programme of training Selection to medical school
Amplitude tests	UKCAT BMAT	Measure performance across a range of mental abilities	Selection to medical school
Orals and vivas			
FRCS MRCS		Assessment of knowledge by face-to-face interaction between the assessor and the candidate	Certification
Observational assessments of clinical competence			
Work based assessments (DOPS, mini-CEX, CBD, PBA, OSATS)		Assessment of clinical competence with direct observation of the candidate performance in a real clinical or simulated context	Progression in a programme of training
OSCE			Certification
Interviews		Depending on the purpose they combine different elements, such as skills stations, CV stations, and assessment centres.	Selection to a programme of training Selection to medical school
Portfolios		Record of achievements and experiences	Progression in a programme of training

MCQ, multiple choice question; EMQ, extended matching question; SAQ, short-answer question; GCSE, General Certificate of Secondary Education; CV, curriculum vitae; UKCAT, UK Clinical Aptitude Test; BMAT, Biomedical Admissions Test; FRCS, Fellow of the Royal College of Surgeons; MRCS, Member of the Royal College of Surgeons; DOPS, direct observation of procedural skills; mini-CEX, mini-clinical evaluation exercise; CBD, case-based discussion; PBA, procedure-based assessments; OSATS, objective structured assessment of technical skills; OSCE, objective structured clinical examination.

mains, an appropriate number of stations, and using different examiners for each station [37]. Research has shown that global rating scores increase the construct validity of OSCEs, assessing expertise better than detailed checklists. Care should also be taken not to sacrifice validity by reducing the time needed for the assessment of clinical skills in an attempt to increase reliability by increasing the number of stations [3].

CONCLUSION

In the long journey of surgical training, the same assessment formats are frequently being used for selection into a programme of study, job selection, progression, and certification (Table 1). These assessment methods must ensure that the appropriate candidates are selected into a programme of study or job and must guarantee public safety by regulating the progression of surgical trainees and the certification of trained surgeons. Although written tests, such as MCQs and EMQs, have been proven to have appropriate validity and reliability for the purposes of selection into medical school, their psychometric properties have not been examine for certification purposes, such as MRCS and FRCS. Also, although assessments of clinical competence have been proven very reliable and valid in the context of medical school final exams (OSCEs) and progression into training (WBAs), their psychometric properties need to be examined in the context of their emerging role as tools in job selection. The psychometric properties of the various assessment methods are different for each purpose, and because of the significance these assessments have for trainees and patients, their reliability and validity should be examined thoroughly in every context where the assessment method is being used.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

1. van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J. Objective assessment of technical surgical skills. Br J Surg. 2010;97:972-987.
2. Beard JD. Assessment of surgical skills of trainees in the UK. Ann R Coll Surg Engl. 2008;90:282-285.
3. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet. 2001;357:945-949.
4. Searle J, McHarg J. Selection for medical school: just pick the right students and the rest is easy! Med Educ. 2003;37:458-463.
5. Benbassat J, Baumal R. Uncertainties in the selection of applicants for medical school. Adv Health Sci Educ Theory Pract. 2007; 12:509-521.

6. 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;332:1005-1009.
7. Randall R, Davies H, Patterson F, Farrell K. Selecting doctors for postgraduate training in paediatrics using a competency based assessment centre. Arch Dis Child. 2006;91:444-448.
8. Patterson F, Ferguson E, Thomas S. Using job analysis to identify core and specific competencies: implications for selection and recruitment. Med Educ. 2008;42:1195-1204.
9. 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 non-technical skills for recruitment to training in anaesthesia. Br J Anaesth. 2010;105:603-609.
10. Davies H, Archer J, Southgate L, Norcini J. Initial evaluation of the first year of the Foundation Assessment Programme. Med Educ. 2009;43:74-81.
11. Welchman SA. Educating the surgeons of the future: the successes, pitfalls and principles of the ISCP. Bull R Coll Surg Eng. 2012; 94(2):1-3.
12. Pelgrim EA, Kramer AW, Mokkink HG, van den Elsen L, Grol RP, van der Vleuten CP. In-training assessment using direct observation of single-patient encounters: a literature review. Adv Health Sci Educ Theory Pract. 2011;16:131-142.
13. McKee RF. The intercollegiate surgical curriculum programme (ISCP). Surgery (Oxford). 2008;26:411-416.
14. Mitchell C, Bhat S, Herbert A, Baker P. Workplace-based assessments of junior doctors: do scores predict training difficulties? Med Educ. 2011;45:1190-1198.
15. Driessen E, van Tartwijk J, van der Vleuten C, Wass V. Portfolios in medical education: why do they meet with mixed success? A systematic review. Med Educ. 2007;41:1224-1233.
16. Tochel C, Haig A, Hesketh A, Cadzow A, Beggs K, Colthart I, Peacock H. The effectiveness of portfolios for post-graduate assessment and education: BEME Guide No 12. Med Teach. 2009;31: 299-318.
17. Hutchinson L, Aitken P, Hayes T. Are medical postgraduate certification processes valid? A systematic review of the published evidence. Med Educ. 2002;36:73-91.
18. Pandey VA, Wolfe JH, Lindahl AK, Rauwerda JA, Bergqvist D; European Board of Vascular Surgery. Validity of an exam assessment in surgical skill: EBSQ-VASC pilot study. Eur J Vasc Endovasc Surg. 2004;27:341-348.
19. Iqbal IZ, Naqvi S, Abeysundara L, Narula AA. The value of oral assessments: a review. Bull R Coll Surg Eng. 2010;92(7):1-6.
20. Ward D, Parker M. The new MRCS: changes from 2008. Bull R Coll Surg Eng. 2009;91:88-90.
21. McGaghie WC, Cohen ER, Wayne DB. Are United States Medical Licensing Exam Step 1 and 2 scores valid measures for postgraduate medical residency selection decisions? Acad Med. 2011;86:48-52.
22. 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;9:417-420.
23. 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;43:50-57.
24. Oosterveld P, ten Cate O. Generalizability of a study sample assessment procedure for entrance selection for medical school. Med Teach. 2004;26:635-639.
25. Turner R, Nicholson S. Can the UK Clinical Aptitude Test (UKCAT) select suitable candidates for interview? Med Educ. 2011;45:1041-1047.
26. 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;33:53-57.
27. Emery JL, Bell JF. The predictive validity of the BioMedical Admissions Test for pre-clinical examination performance. Med Educ. 2009;43:557-564.
28. Lynch B, Mackenzie R, Dowell J, Cleland J, Prescott G. Does the UKCAT predict Year 1 performance in medical school? Med Educ. 2009;43:1203-1209.
29. Davis MH, Karunathilake I. The place of the oral examination in today's assessment systems. Med Teach. 2005;27:294-297.
30. Bandiera G, Regehr G. Reliability of a structured interview scoring instrument for a Canadian postgraduate emergency medicine training program. Acad Emerg Med. 2004;11:27-32.
31. Dore KL, Kreuger S, Ladhani M, Rolfson D, Kurtz D, Kulasegaram K, Cullimore AJ, Norman GR, Eva KW, Bates S, Reiter HI. The reliability and acceptability of the Multiple Mini-Interview as a selection instrument for postgraduate admissions. Acad Med. 2010;85(10 Suppl):S60-S63.
32. Humphrey S, Dowson S, Wall D, Diwakar V, Goodyear HM. Multiple mini-interviews: opinions of candidates and interviewers. Med Educ. 2008;42:207-213.
33. Urlings-Strop LC, Stijnen T, Themmen AP, Splinter TA. Selection of medical students: a controlled experiment. Med Educ. 2009; 43:175-183.
34. Patterson F, Ferguson E, Norfolk T, Lane P. A new selection system to recruit general practice registrars: preliminary findings from a validation study. BMJ. 2005;330:711-714.
35. Roberts C, Newble DI, O'Rourke AJ. Portfolio-based assessments in medical education: are they valid and reliable for summative purposes? Med Educ. 2002;36:899-900.
36. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. Med Teach. 2007;29:855-871.
37. van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. Med Educ. 2005;39:309-317.