



Establishing the predictive validity of the intercollegiate membership of the Royal Colleges of surgeons written examination: MRCS part B

Ricky Ellis ^{a,b,*}, Jennifer Cleland ^c, Duncan SG. Scrimgeour ^{a,d},
Amanda J. Lee ^e, John Hines ^f, Peter A. Brennan ^g

^a Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, United Kingdom

^b Urology Department, Nottingham University Hospitals, Nottingham, United Kingdom

^c Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

^d Department of Colorectal Surgery, Aberdeen Royal Infirmary, Aberdeen, United Kingdom

^e Medical Statistics Team, Institute of Applied Health Sciences, University of Aberdeen, United Kingdom

^f University College Hospital London, United Kingdom

^g Department of Maxillo-Facial Surgery, Queen Alexandra Hospital, Portsmouth, United Kingdom

ARTICLE INFO

Article history:

Received 14 April 2023

Received in revised form

12 July 2023

Accepted 13 July 2023

Available online 28 July 2023

Keywords:

Surgery

Medical education & training

ABSTRACT

The Intercollegiate Membership of the Royal Colleges of Surgeons (MRCS) is a high-stakes postgraduate examination taken by thousands of surgical trainees worldwide every year. The MRCS is a challenging assessment, highly regarded by surgical training programmes and valued as a gatekeeper to the surgical profession. The examination is taken at considerable personal, social and financial cost to surgical trainees, and failure has significant implications for career progression. Given the value placed on MRCS, it must be a reliable and valid assessment of the knowledge and skills of early-career surgeons.

Our first article 'Establishing the Predictive Validity of the Intercollegiate Membership of the Royal Colleges of Surgeons Written Examination: MRCS Part A' discussed the principles of assessment reliability and validity and outlined the mounting evidence supporting the predictive validity of the MRCS Part A (the multiple-choice questionnaire component of the examination). This, the second article in the series discusses six recently published studies investigating the predictive validity of the MRCS Part B (the clinical component of the examination).

All national longitudinal cohort studies reviewed have demonstrated significant correlations between MRCS Part B and other assessments taken during the UK surgical training pathway, supporting the predictive validity of MRCS Part B. This review will be of interest to trainees, trainers and Royal Colleges given the value placed on the examination by surgical training programmes.

© 2023 The Author(s). Published by Elsevier Ltd on behalf of Royal College of Surgeons of Edinburgh (Scottish charity number SC005317) and Royal College of Surgeons in Ireland.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

The Intercollegiate Membership of the Royal Colleges of Surgeons (MRCS) examination is a postgraduate assessment attempted by thousands of surgical trainees worldwide every year.¹ It comprises two parts; Part A (a written, multiple-

choice questionnaire format examination) and Part B (the Objective Structured Clinical Examination (OSCE)). Successful completion of both Parts A and B of the MRCS are a prerequisite for progression into surgical higher specialist training (HST) programmes in the United Kingdom (UK).¹ As such, the MRCS is a key assessment, functioning as a gatekeeper to the

* Corresponding author. Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, AB25 2ZD, United Kingdom.

E-mail addresses: Rickyellis@nhs.net (R. Ellis), jennifer.cleland@ntu.edu.sg (J. Cleland), duncan.scrimgeour@nhs.scot (D.S.G. Scrimgeour), a.j.lee@abdn.ac.uk (A.J. Lee), Jhines@rcseng.ac.uk (J. Hines), peter.brennan@porthosp.nhs.uk (P.A. Brennan).
<https://doi.org/10.1016/j.surge.2023.07.003>

1479-666X/© 2023 The Author(s). Published by Elsevier Ltd on behalf of Royal College of Surgeons of Edinburgh (Scottish charity number SC005317) and Royal College of Surgeons in Ireland. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

surgical profession and as a safeguard for patients, ensuring that surgeons entering HST have met a universally agreed standard.^{2,3} To succeed at MRCS, trainees must, therefore, demonstrate that they have acquired the knowledge, skills and attributes expected of them by the completion of Core Surgical Training (CST).⁴

Given the value placed on MRCS by surgical training programmes and the considerable personal, social and financial sacrifices made by trainees while attempting the examination,⁵ it is essential that the MRCS, like any other high-stakes examination, is reliable and valid.⁶

In the first article of this series,⁷ we discussed the principles of assessment reliability (a measure of the reproducibility of the examination and its results^{8,9}). The quantitative assessment of MRCS reliability measures is published annually by the Intercollegiate Committee for Basic Surgical Examinations (ICBSE)¹⁰ and ranges from 0.59 to 0.88 (using Cronbach's alpha¹¹) for Part B, which is comparable to the reliability measures of other high-stakes postgraduate medical examinations.¹²

Unlike reliability, the validity of an examination is much more difficult to assess. The first of our two papers in this series presented the mounting evidence in support of the predictive validity of MRCS Part A. Until recently, little was known about the predictive validity of the MRCS Part B. Yet if the MRCS Part B is to continue to be used as a key gatekeeper for the surgical profession, it must also demonstrate validity to be deemed fit for purpose. This article discusses the principles of examination validity and summarises the findings of recent studies undertaken to investigate the predictive validity of the MRCS Part B clinical examination.

MRCS part B

MRCS Part B comprises approximately 18 OSCE stations, each taking 9 min to complete, with 2 min between each to enable candidates to move to the next station and read the scenario, while allowing examiners to ensure marks have been recorded. Two preparation stations are added before each communication skills testing station, giving candidates time to read and assimilate clinical information before discussing the clinical case. Two rest stations are also included in the circuit bringing the total number of stations to 22.⁴ The number of stations has recently been reduced to 17, although this was bought into place in October 2021, after the studies featured in this review were conducted.¹³

MRCS Part B stations test two broad content areas (BCA); 'Knowledge' and 'Skills'. Eight stations test 'Knowledge' and ten stations test 'Skills'. The 'Knowledge' BCA includes anatomy, surgical pathology, applied surgical science and critical care. The 'Skills' BCA includes communication, clinical and procedural skills. OSCE stations may take a variety of forms, including anatomy demonstration and discussion, mock clinical consultations, viva voce testing of knowledge and the demonstration of clinical skills.⁴ Performance at each station is quantified by the award of a score for each BCA and also a judgement of the candidate's performance on the whole (pass, borderline or fail). Pass marks for each station are determined using borderline regression methodology (commonly used by other institutions delivering postgraduate examinations).⁴

Candidates must pass MRCS Part A to be eligible to attempt Part B. Upon completing both parts of the examination, candidates are awarded a postgraduate Diploma and are granted permission to use the prestigious MRCS post-nominals.

Validity measures

Unlike the relatively simple measurement of the reliability of an assessment, measuring the validity is more challenging.¹⁴ Validity describes whether or not an assessment measures what it intends to measure. The first article in this series described the various components that constitute the validity of an examination.⁷ In short, these include 'face validity' (the examination tests what it intends to test), 'content validity' (the examination tests knowledge of the curriculum) and 'predictive validity' (the ability of a test to predict future outcomes). Careful alignment between examination questions and both the Intercollegiate Surgical Curriculum Programme¹⁵ and the General Medical Council (GMC) framework on Generic Professional Capabilities¹⁶ ensures the content validity of the MRCS Part B. The careful selection of OSCEs from a large pool of highly regulated stations ensures that Part B examines appropriately across the entire curriculum to ensure the face validity of the examination.

The predictive validity of an assessment (its ability to predict future outcomes) is even more challenging to assess. The GMC has confirmed that in the absence of a gold standard with which to compare medical examinations, one way of establishing the predictive validity of an assessment is to 'establish the strength of the relationships between similar assessments'.¹⁷ As such, performance in one test should predict the performance in a similar test taken during trainees' educational careers and correlate with performance in previous tests.

Until recently, the predictive validity of the MRCS was largely untested. This was an important research question as success on other postgraduate medical examinations in the UK and elsewhere has been found to predict future clinical performance measures.^{18–21} For example, the predictive validity of the Membership of the Royal College of Physicians (MRCP) has been established using comparisons with each of its three components²² in addition to A-Levels, medical school admission test scores, medical school performance, and performance at MRCGP and the Professional and Linguistic Assessment Board (PLAB) test.^{23–27}

Methods

To address this gap in the literature, several large longitudinal cohort studies were undertaken by two ICBSE Research Fellows and their associated research teams to investigate the predictive validity of the MRCS.²⁸ This article summarises the findings of these studies in the context of establishing the predictive validity of the MRCS Part B clinical examination. Each study used a combination of univariate analyses, Pearson correlation coefficients and logistic regression modelling to assess the relationship between each examination and MRCS Part B pass/fail outcomes. First attempt results were commonly used as these are known to be the

best predictor of future performance in postgraduate medical examinations.²²

Results

The six longitudinal cohort studies and their combined utility in the context of assessing the predictive validity of the MRCS Part B examination are presented in Table 1.^{29–34}

School

The first study by Ellis et al.²⁹ demonstrated the challenge of investigating the predictive validity of postgraduate clinical assessments. Unlike formal written examinations, there are few high-stakes clinical examinations taken throughout the training pathway with which to compare the MRCS Part B. In this longitudinal cohort study, Ellis et al.²⁹ compared performance at A-Levels (high-school exit examinations) and medical school admissions test results with MRCS Part B outcomes. Candidates who passed MRCS Part B at the first attempt had higher A-Level and admission test scores when compared to those who failed at the first attempt, although this did not reach statistical significance. A-Levels (Odds Ratio [OR] 1.07, 95% Confidence Interval [CI] 1.03 to 1.11), Biomedical Admissions Test (BMAT) scores (OR 1.16, 95% CI 1.00 to 1.34) and specifically the BMAT Aptitude and Skills subtest scores (OR 1.31, 95% CI 1.04 to 1.67) were the only statistically significant independent predictors of MRCS Part

B performance after adjusting for sociodemographic predictors of success. However, the predictive value of these tests diminished after A* A-Level grades were introduced in 2010 to combat A-Level grade inflation. It is challenging to apply these results to the predictive validity of the MRCS Part B examination, given the comparison between written assessments of school-level knowledge and aptitude (admissions tests) with a later clinical examination of knowledge and skill. These assessments are unlikely to be described as 'similar' according to the GMC's assessment guidance.¹⁷ The subsequent comparison of performance in written and clinical assessments at medical school by Ellis et al.³⁰ was, therefore, more applicable to the assessment of the predictive validity of Part B.

Medical school

On completion of medical school, all graduates are ranked across the UK for selection to Foundation Programme (FP) training posts. National rankings use the combined performance scores from a Situational Judgement Test (FP-SJT) and the Educational Performance Measure (EPM).³⁵ The EPM is a score comprised of; a student's performance decile within each medical school, additional points awarded for peer-reviewed publications and additional points awarded for other degree-level qualifications.³⁶ The EPM decile is awarded based on performance in a number of different assessments (written and clinical) of students' knowledge and skills compared to their peers.

Table 1 – Primary outcomes measured in each of the studies by the ICBSE Research group. Presented in chronological order of when each assessment is taken in the UK surgical training pathway.

Study	Title	Assessments Compared to MRCS Part B	Number of candidates
Ellis et al. 2021 <i>Postgraduate Medical Journal</i> . ²⁹	Performance at medical school selection correlates with success in Part A of the Intercollegiate Membership of the Royal Colleges of Surgeons (MRCS) Examination	A-Levels University Clinical Aptitude Test (UCAT) Biomedical Admissions Test (BMAT) Graduate Medical School Admissions Test (GAMSAT)	5690
Ellis et al. 2021 <i>BMJ Open</i> . ³⁰	Does performance at medical school predict success at the Intercollegiate Membership of the Royal Colleges of Surgeons (MRCS) examination? A retrospective cohort study	Educational Performance Measure (EPM) Foundation Programme Situational Judgement Test (FP-SJT)	755
Scrimgeour et al., 2018 <i>The Surgeon</i> . ³¹	Which factors predict success in the mandatory UK postgraduate surgical exam: The Intercollegiate Membership of the Royal Colleges of Surgeons (MRCS)?	Membership of the Royal Colleges of Surgeons (MRCS) Part A	4310
Scrimgeour et al., 2017 <i>BJS Open</i> . ³²	Impact of performance in a mandatory postgraduate surgical examination on selection into specialty training: Performance in a postgraduate surgical examination and selection into specialty training	HST (ST3) National Selection Score for General and Vascular surgery	774
Scrimgeour et al., 2018 <i>Annals of the Royal College of Surgeons of England</i> . ³³	Does the Intercollegiate Membership of the Royal College of Surgeons (MRCS) examination predict 'on-the-job' performance during UK higher specialty surgical training?	HST Annual Review of Competence Progression (ARCP) outcomes	2570
Scrimgeour et al., 2019 <i>BJS Open</i> . ³⁴	Prediction of success at UK Specialty Board Examinations using the mandatory postgraduate UK surgical examination	Fellowship of the Royal Colleges of Surgeons (FRCS)	854

Ellis et al.³⁰ found that candidates who passed MRCS Part B at their first attempt had higher total EPM scores compared with those who failed at the first attempt (mean score 43.3 (Standard Deviation [SD] 3.6) vs 41.1 (SD 3.4), $p < 0.001$). The likelihood of passing MRCS Part B increased by 23% (OR 1.23, 95% CI 1.14 to 1.32) for every additional EPM decile achieved. This data supports the predictive validity of MRCS Part B. Unfortunately, the EPM does not allow data from medical school written and clinical examinations to be scrutinised separately, so it was not possible to examine whether clinical scores are more strongly correlated with Part B performance.

FP-SJT performance was not an independent predictor of Part B success.³⁰ Interestingly, given that there is no statistically significant association between FP-SJT performance and later disciplinary action by the GMC, the predictive validity of the FP-SJT itself has been brought into question.³⁷

MRCS part A

Scrimgeour et al.³¹ found a statistically significant moderate correlation between MRCS Part A scores and Part B scores at the first attempt ($r = 0.41$, $p < 0.001$). On multivariate regression analyses, Part A score was found to independently predict Part B success (OR 1.10, 95% CI 1.09 to 1.12), strongly supporting the predictive validity of the examination. In addition, the odds of passing MRCS Part B at the first attempt decreased by 30% for every additional required to pass MRCS Part A (OR 0.70, 95% CI 0.61 to 0.81).

National selection

Recruitment for surgical HST posts is competitive, and each surgical specialty runs a national selection process. Candidates rank their preferred HST deaneries (training locations) and posts are allocated in order of candidates' national selection scores. National selection scores are calculated from performance at interview, assessment of academic portfolios and performance on OSCE stations used to measure candidates' knowledge and skill. HST national selection is, therefore, a 'similar' assessment to MRCS Part B.

Scrimgeour et al.³² found statistically significant moderate correlations between MRCS Part B first and passing attempt scores and national selection scores ($r = 0.38$ and $r = 0.30$ respectively, $p < 0.001$) for General and Vascular surgery. Part B score on passing, and the number of attempts required to pass the examination were both found to be independent predictors of national selection score (change in R^2 of 0.10 and 0.07 respectively, $p < 0.001$). This paper uses data from General and Vascular surgery national selection only, and data from national selection for other surgical specialities are currently being analysed. Despite this, the current data strongly supports the predictive validity of the MRCS Part B.

Annual review of competence progression (ARCP)

Progression through UK surgical HST is regulated using the Annual Review of Competence Progression (ARCP) process. This is a summative process during which the ARCP assessment panel (typically consisting of the Training Programme Director and several senior trainers) examine each trainees'

portfolio of evidence that includes work-based assessments, operative logbooks, supervisor reports and academic achievements (including audit and quality improvement projects, publications, presentations at meetings and leadership roles). The panel uses nationally established curricula to determine whether each trainee is competent to progress to the next level of their training.³⁸

Scrimgeour et al.³³ assessed whether MRCS Part B scores were associated with HST Annual Review of Competence Progression (ARCP) outcomes. Trainees that passed MRCS Part B at the first attempt were significantly more likely to achieve 'Satisfactory' ARCP outcomes than those who required two or more attempts to pass (63.1% vs 57.1% respectively, $p < 0.001$). MRCS Part B score and the number of attempts taken to pass Part B were both found to be independent predictors of 'Unsatisfactory' ARCP outcomes (OR 0.98, 95% CI 0.97 to 1.00 and OR 1.50, 95% CI 1.16 to 1.94 respectively). These data strongly support the predictive validity of the MRCS Part B, confirming that Part B successfully predicts 'on-the-job' performance in HST.

FRCS

The Fellowship of the Royal Colleges of Surgeons (FRCS examination) is a high-stakes postgraduate specialty assessment taken during HST. FRCS comprises Section 1, a written examination, and Section 2, an OSCE examination.³⁹ Successful completion of both parts is a prerequisite for the award of Certificate of Completion of Training (CCT) in the UK, enabling a surgeon to apply for consultant posts.⁴⁰ As such, FRCS is comparable to MRCS in function and format.

Scrimgeour et al.³⁴ found that MRCS Part B score and the number of attempts required to pass Part B were independent predictors of first-attempt FRCS Section 1 success (OR 1.06, 95% CI 1.03 to 1.09 and OR 1.77, 95% CI 1.08 to 3.00). Surprisingly, MRCS Part B performance was not found to independently predict FRCS Section 2 success in multivariate regression analyses, but there were statistically significant correlations between Part B scores and both FRCS Section 1 ($r = 0.29$, $p < 0.001$) and FRCS Section 2 scores ($r = 0.34$, $p < 0.001$).

Discussion

Main findings

Investigating the predictive validity of a high-stakes postgraduate clinical examination is inherently difficult owing to the lack of 'similar' comparative assessments taken throughout trainees' educational careers. However, the large-scale longitudinal cohort studies recently undertaken to answer this key research question have presented generalisable data that supports the predictive validity of the MRCS Part B examination.^{29–34} Fig. 1 shows the assessments that correlate with MRCS Part B. These data ensure the evidence base for predictive validity of MRCS Part B is commensurate with that of other UK postgraduate medical examinations. Additionally, candidates who pass Part B at the first attempt are more than twice as likely to continue in surgical careers

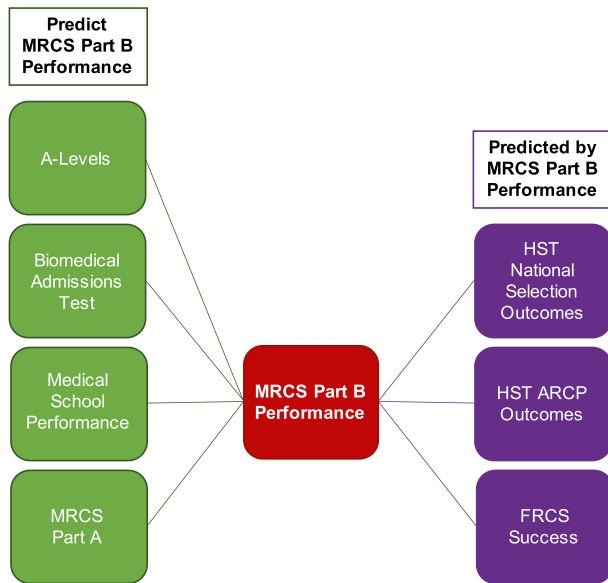


Fig. 1 – Assessments with statistically significant associations to MRCS Part B performance.

than those who fail, many of whom leave surgical training to enter other medical specialties (OR 2.03, 95% CI 1.58 to 2.62).² These findings, in addition to the mounting evidence of the predictive validity of MRCS Part B, support its continued use in surgical training programmes.

Predicting success at FRCS

MRCS Part B performance was not found to independently predict FRCS Section 2 success.³⁴ One possible reason for this may include differences in the aims and assessment methods of each of the two examinations. MRCS is an examination of the knowledge, clinical and procedural skills expected of surgical trainees at the completion of CST which includes the 'Principles of Surgery in General' and a broad range of surgical conditions. Part B is an OSCE examination requiring candidates to demonstrate their ability to take a clinical history, perform a clinical examination or basic surgical skills. On the other hand, FRCS is a specialty board examination, designed to examine the knowledge and skills expected of a newly appointed consultant in their chosen surgical specialty. Section 2 focuses on scenario-based structured interviews examining the candidate's knowledge and the application of an evidence-based management plan, thereby focussing on senior clinical decision-making. This subtle distinction equates to a significant difference in what is being examined (i.e., the ability to progress from general CST into specialty HST vs the ability to be a day-one consultant in a surgical specialty) and may be a contributing factor to MRCS Part B not being a predictor of FRCS Section 2 success. The positive correlation between MRCS Part B and FRCS Section 2 scores suggests a degree of alignment between examinations but this is not strong enough given the differences in aims and methods of each examination to enable trainers to use MRCS Part B scores to predict those at increased risk of failing FRCS Section 2 for early remedial training or additional support.

Implications for practice

It is important to consider the limitations of one-off high-stakes examinations in comparison to more regular assessments of knowledge, skills and competencies within the clinical environment.⁴¹ Does performance at a high-stakes examination capture the breadth of a candidate's clinical knowledge, surgical skill, communication skills in difficult conversations or non-technical skills in emergency scenarios encountered regularly in clinical practice? Does performance at MRCS rely too heavily on peak performance for only two days of a candidate's surgical career?

Recent changes in the surgical curriculum aim to improve the capture of multiple assessments of surgical trainees' skills and knowledge within the workplace on a daily basis using a variety of work-based assessments that are conducted by trainers and reviewed at annual appraisals alongside reports from multiple consultants/trainers within each training unit.³⁸ Given this robust assessment of workplace performance and clinical progression it calls into question whether the stakes of postgraduate examinations should be lowered.

Postgraduate examinations such as MRCS represent an objective assessment of skill and knowledge that may have a place alongside lower-stakes work-based assessments to account for the known challenge of some supervisors 'failing to fail' underperforming trainees.⁴² The studies reviewed in this article also support the predictive validity of the MRCS Part B which is currently used as a gatekeeper to entry into HST and appears to successfully predict the likelihood of continuing in a surgical career.² However, there can be little doubt that high-stakes postgraduate assessments are taken by trainees at great cost in terms of time, money and work-life balance.⁵ We welcome debate by the surgical community and educationalists as to whether there should be a shift in the aims of postgraduate assessment to 'assessment for learning' and away from 'assessment of learning'. If so, whether this could be achieved by changes that include lowering the stakes of examinations, increasing the frequency of lower-stakes examinations (as has been used in the United States⁴³) and improving the granularity of candidate feedback for example.

Future work

Future work to further assess the predictive validity of Part B may reasonably include comparisons with other markers of clinical performance in surgery and surgical training. For example, lower examination scores in other postgraduate examinations such as the MRCP, MRCGP, PLAB, American Board of Internal Medicine certification examinations and the USMLE are associated with an increased likelihood of sanctions and disciplinary action by medical regulators such as the GMC.^{21,44–46} Ellis et al., 2021 investigated whether there was an association between MRCS Part B performance and GMC Fitness to Practice sanctions (FtP). However, the encouragingly small number of FtP sanctions within the large cohort of surgical trainees in the UK (31 sanctions across 11,660 surgeons), prevented any meaningful statistical analyses of these variables.⁴⁷ Future studies may wish to consider comparisons with other clinical outcome measures such as operative outcome data, patient complaints and time to CCT.

Conclusion

Recent large-scale longitudinal cohort studies have demonstrated significant correlations between MRCS Part B and other assessments taken in the UK surgical training pathway. These studies support the predictive validity of MRCS Part B, a key gatekeeping assessment in UK surgical training.

Funding

Royal College of Surgeons of Ireland, Royal College of Physicians and Surgeons of Glasgow, Royal College of Surgeons of England and Royal College of Surgeons of Edinburgh (Award/Grant number is not applicable).

Contributorship

RE wrote the first draft of the manuscript. PB, JC, DS, AL and JH reviewed and edited the manuscript. All authors have approved the final copy.

Declaration of competing interest

None to declare.

Acknowledgements

The authors would like to acknowledge Iain Target at the Royal College of Surgeons of England, and Gregory Ayre from the Intercollegiate Committee for Basic Surgical Examinations for their support during this project.

Abbreviations

ARCP	Annual Review of Competence Progression
BMAT	Biomedical Admissions Test
CCT	Certificate of Completion of Training
CST	Core Surgical Training
EPM	Educational Performance Measure
FP	Foundation Training Programme
FP-SJT	Foundation Programme Situational Judgement Test (SJT)
FRCS	Fellowship of the Royal Colleges of Surgeons
FtP	Fitness to Practice
GMC	General Medical Council
ICBSE	Intercollegiate Committee for Basic Surgical Examinations
MCQ	Multiple Choice Questionnaire
MRCP	Membership of the Royal College of Physicians
MRCGP	Membership of the Royal College of General Practitioners
MRCS	Membership of the Royal Colleges of Surgeons
OSCE	Objective Structured Clinical Examination
PLAB	Professional and Linguistic Assessment Board Test

UK	United Kingdom
USMLE	United States Medical Licensing Examination

REFERENCES

- Brennan PA, Sherman KP. The MRCS examination—an update on the latest facts and figures. *Br J Oral Maxillofac Surg* 2014;52(10):881–3.
- Ellis R, Cleland J, Scrimgeour D, Lee A, Brennan P. Does the MRCS exam fulfil its purpose in surgical professions? *Bull Roy Coll Surg Engl* 2021;103(7):344–50.
- Nagy P. The three roles of assessment: gatekeeping, accountability, and instructional diagnosis. *Can J Educ Rev Can Léducation*. 2000;25(4):262.
- Guide to the intercollegiate membership of the royal College of surgeons examinations [internet]. Intercollegiate committee for basic surgical examinations. 2018 [cited 2020 Sep 23]. Available from: <https://www.intercollegiatemrcsexams.org.uk/-/media/files/imrcs/mrcs/mrcs-candidate-guidance/candidate-guide-to-mrcs-examination-january-2019-clean.pdf>.
- Ellis R, Scrimgeour DSG, Brennan PA. The personal cost of postgraduate medical exams: are we asking too much of trainees? *Br Med J [Internet]* 2021 [cited 2021 May 21];thebmjopinion. Available from: <https://blogs.bmj.com/bmj/2021/02/02/the-personal-cost-of-postgraduate-medical-exams-are-we-asking-too-much-of-trainees/>.
- The standards for educational and psychological testing*. Washington, DC, U.S: American Educational Research Association, American Psychological Association, National Council on Measurement in Education and Joint Committee on Standards for Educational and Psychological Testing; 2014.
- Ellis R, Cleland J, Scrimgeour DSG, Lee AJ, Hines J, Brennan PA. Establishing the predictive validity of the intercollegiate membership of the Royal Colleges of surgeons written examination: MRCS Part A. *The Surgeon* 2023. <https://doi.org/10.1016/j.surge.2023.07.004>. in press. [Online ahead of print].
- Hutchinson LW. Assessing medical competence: finding the right answers. *Clin Teach* 2004;1(1):14–8.
- Downing SM. Reliability: on the reproducibility of assessment data. *Med Educ* 2004;38(9):1006–12.
- Intercollegiate Committee for Basic Surgical Examinations 2018/19 Annual Report [Internet]. *Intercollegiate committee for basic surgical examinations*. 2019 [cited 2020 Sep 23]. Available from: <https://www.intercollegiatemrcsexams.org.uk/-/media/files/imrcs/policies-and-reports/icbse-annual-report-201819-final.docx>.
- Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16(3):297–334.
- Hutchinson L, Aitken P, Hayes T. Are medical postgraduate certification processes valid? A systematic review of the published evidence. *Med Educ* 2002;36(1):73–91.
- MRCS Part B october 2021 confirmation of format and delivery [internet]. Intercollegiate Committee for Basic Surgical Examinations; 2021 [cited 2021 Nov 23]. Available from: <https://www.intercollegiatemrcsexams.org.uk/announcements/mrcs-part-b-october-2021-confirmation-of-format-and-delivery/>.
- Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ* 2003;37(9):830–7.
- Intercollegiate Surgical Curriculum Programme (ISCP) [Internet]. Joint Committee on Surgical Training; [cited 2020 Sep 23]. Available from: <https://www.iscp.ac.uk/curriculum/surgical/curriculum.aspx>.
- Excellence by design: standards for postgraduate curricula [internet]. General Medical Council; 2017 [cited 2020 Sep 23]. Available from: <https://www.gmc-uk.org/-/media/>

- documents/Excellence_by_design__standards_for_postgraduate_curricula_0517.pdf_70436125.pdf.
17. Designing and maintaining postgraduate assessment programmes. General Medical Council. [Internet]. [cited 2020 Apr 28]. Available from: https://www.gmc-uk.org/-/media/documents/designing-and-maintaining-postgraduate-assessment-programmes-0517_pdf-70434370.pdf.
 18. Wenghofer E, Klass D, Abrahamowicz M, Dauphinee D, Jacques A, Smee S, et al. Doctor scores on national qualifying examinations predict quality of care in future practice. *Med Educ* 2009;43(12):1166–73.
 19. Tamblyn R, Abrahamowicz M, Dauphinee D, Wenghofer E, Jacques A, Klass D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. *JAMA* 2007;298(9):993–1001.
 20. Norcini JJ, Boulet JR, Opalek A, Dauphinee WD. The relationship between licensing examination performance and the outcomes of care by international medical school graduates. *Acad Med J Assoc Am Med Coll* 2014;89(8):1157–62.
 21. Wakeford R, Ludka K, Woolf K, McManus IC. Fitness to practise sanctions in UK doctors are predicted by poor performance at MRCP(UK) assessments: data linkage study. *BMC Med* 2018;16(1):230.
 22. McManus I, Ludka K. Resitting a high-stakes postgraduate medical examination on multiple occasions: nonlinear multilevel modelling of performance in the MRCP(UK) examinations. *BMC Med* 2012;10(1):60.
 23. Wakeford R, Denney M, Ludka-Stempien K, Dacre J, McManus IC. Cross-comparison of MRCP(UK) & MRCP(UK) in a database linkage study of 2,284 candidates taking both examinations: assessment of validity and differential performance by ethnicity. *BMC Med Educ* 2015;15(1):1.
 24. McManus IC, Wakeford R. PLAB and UK graduates' performance on MRCP(UK) and MRCP(UK) examinations: data linkage study. *BMJ* 2014;348:g2621.
 25. McManus I, Woolf K, Dacre J, Paice E, Dewberry C. The Academic Backbone: longitudinal continuities in educational achievement from secondary school and medical school to MRCP(UK) and the specialist register in UK medical students and doctors. *BMC Med* 2013;11(1):242.
 26. McManus IC, Harborne AC, Horsfall HL, Joseph T, Smith DT, Marshall-Andon T, et al. Exploring UK medical school differences: the MedDifs study of selection, teaching, student and F1 perceptions, postgraduate outcomes and fitness to practise. *BMC Med* 2020;18(1):136.
 27. McManus IC, Woolf K, Harrison D, Tiffin PA, Paton LW, Cheung KYF, et al. Predictive validity of A-level grades and teacher-predicted grades in UK medical school applicants: a retrospective analysis of administrative data in a time of COVID-19. *BMJ Open* 2021;11(12):e047354.
 28. Intercollegiate Committee for Basic Surgical Examinations 2017/18 Annual Report [Internet]. *Intercollegiate committee for basic surgical examinations*. 2018 [cited 2020 Sep 30]. Available from: <https://www.intercollegiatemrcsexams.org.uk/-/media/files/imrcs/policies-and-reports/icbseannual-report201718final.pdf>.
 29. Ellis R, Brennan P, Scrimgeour DS, Lee AJ, Cleland J. Performance at medical school selection correlates with success in Part A of the intercollegiate Membership of the Royal College of Surgeons (MRCS) examination. *Postgrad Med* 2021;98(1161). [postgradmedj-2021-139748](https://doi.org/10.1136/postgradmedj-2021-139748).
 30. Ellis R, Scrimgeour DSG, Brennan PA, Lee AJ, Cleland J. Does performance at medical school predict success at the Intercollegiate Membership of the Royal College of Surgeons (MRCS) examination? A retrospective cohort study. *BMJ Open* 2021;11(8):e046615.
 31. Scrimgeour DSG, Cleland J, Lee AJ, Brennan PA. Which factors predict success in the mandatory UK postgraduate surgical exam: the Intercollegiate Membership of the Royal College of Surgeons (MRCS)? *Surg J R Coll Surg Edinb Irel* 2018;16(4):220–6.
 32. Scrimgeour DSG, Cleland J, Lee AJ, Griffiths G, McKinley AJ, Marx C, et al. Impact of performance in a mandatory postgraduate surgical examination on selection into specialty training: performance in a postgraduate surgical examination and selection into specialty training. *BJS Open* 2017;1(3):67–74.
 33. Scrimgeour D, Brennan P, Griffiths G, Lee A, Smith F, Cleland J. Does the Intercollegiate Membership of the Royal College of Surgeons (MRCS) examination predict 'on-the-job' performance during UK higher specialty surgical training? *Ann R Coll Surg Engl* 2018;100(8):669–75.
 34. Scrimgeour DSG, Cleland J, Lee AJ, Brennan PA. Prediction of success at UK Specialty Board Examinations using the mandatory postgraduate UK surgical examination. *BJS Open* 2019;3(6):865–71.
 35. [Internet] *UKFP 2021 applicants' handbook*. UK Foundation Programme; 2020 [cited 2021 Jan 28]. Available from: <https://foundationprogramme.nhs.uk>.
 36. [Internet] *Educational performance measure (EPM) 2020 framework*. UK Foundation Programme; 2020 [cited 2020 May 19]. Available from: <https://foundationprogramme.nhs.uk/resources/>.
 37. Sam AH, Bala L, Westacott RJ, Brown C. Is academic attainment or situational judgment test performance in medical school associated with the likelihood of disciplinary action? A national retrospective cohort study. *Acad Med* 2021;96(10):1467–75.
 38. *The intercollegiate surgical curriculum [internet]*. 2021 [cited 2021 Nov 15]. Available from: <https://www.iscp.ac.uk/iscp/curriculum-2021/>.
 39. Joint Committee on Intercollegiate Examinations [Internet]. [cited 2020 Sep 29]. Available from: <https://www.jcie.org.uk/>.
 40. Joint Committee on Surgical Training [Internet]. [cited 2020 Sep 29]. Certification Guidelines and Checklists. Available from: <https://www.jcst.org/quality-assurance/certification-guidelines-and-checklists/>.
 41. Schuwirth LWT, Van der Vleuten CPM. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach* 2011;33(6):478–85.
 42. Yepes-Rios M, Dudek N, Duboyce R, Curtis J, Allard RJ, Varpio L. The failure to fail underperforming trainees in health professions education: a BEME systematic review: BEME Guide No. 42. *Med Teach* 2016;38(11):1092–9.
 43. Cassidy DJ, Chakraborty S, Panda N, McKinley SK, Mansur A, Hamdi I, et al. The surgical knowledge 'growth curve': predicting ABSITE scores and identifying 'at-risk' residents. *J Surg Educ* 2021;78(1):50–9.
 44. Tiffin PA, Paton LW, Mwandigha LM, McLachlan JC, Illing J. Predicting fitness to practise events in international medical graduates who registered as UK doctors via the Professional and Linguistic Assessments Board (PLAB) system: a national cohort study. *BMC Med* 2017;15(1):66.
 45. Papadakis MA, Arnold GK, Blank LL, Holmboe ES, Lipner RS. Performance during internal medicine residency training and subsequent disciplinary action by state licensing boards. *Ann Intern Med* 2008;148(11):869–76.
 46. Cuddy MM, Young A, Gelman A, Swanson DB, Johnson DA, Dillon GF, et al. Exploring the relationships between USMLE performance and disciplinary action in practice: a validity study of score inferences from a licensure examination. *Acad Med J Assoc Am Med Coll* 2017;92(12):1780–5.
 47. Ellis R, Cleland J, Scrimgeour DSG, Lee AJ, Brennan PA. A cross-sectional study examining the association between MRCS performance and surgeons receiving sanctions against their medical registration. *Surgeon* 2021;20(4):211–5.