Intercollegiate Committee for Basic Surgical Examinations

2009/10 ANNUAL REPORT

The Membership Examination of the Surgical Royal Colleges of Great Britain

MRCS

The Diploma of Otolaryngology Head & Neck Surgery

DO-HNS

July 2010

CONTENTS   PAGE

The Intercollegiate Committee for Basic Surgical Examinations (ICBSE) would welcome comments on this Annual Report and ways in which it can be improved in future years. If you have comments on this Report please send them to: The Chairman, ICBSE, c/o dfrome@icbse.org.uk

1. Introduction
This is the third Annual Report of the Intercollegiate Committee for Basic Surgical Examinations (ICBSE) and covers the period August 2009 to July 2010.

The purpose of the Annual Report is to provide a definitive source of information about the Membership Examination of the Surgical Royal Colleges of Great Britain (MRCS) and the Diploma of Otolaryngology Head & Neck Surgery (DO-HNS) for all interested stakeholders including candidates, trainers, Assigned Education Supervisors and the public.

The structure, standard and quality assurance of the MRCS and DO-HNS examinations are the responsibility of the ICBSE which has a number of specialist subgroups each responsible for a different aspect of the examination.

2. The MRCS examination: purpose and structure
The Membership Examination of the Surgical Royal Colleges of Great Britain (MRCS) is designed for candidates in the generality part of their specialty training. It is a crucial milestone which must be achieved if trainees are to progress to specialty surgical training as defined by the nine surgical Specialty Advisory Committees (SACs). The purpose of the MRCS is to determine that trainees have acquired the knowledge, skills and attributes required for the completion of core training in surgery and, for trainees following the Intercollegiate Surgical Curriculum Programme, to determine their ability to progress to higher specialist training in surgery.

Having achieved the intended outcomes of the curriculum it is anticipated that the surgical trainee will be able to perform as a member of the team caring for surgical patients. He or she will be able to receive patients as emergencies, review patients in clinics and initiate management and diagnostic processes based on a reasonable differential diagnosis. He or she will be able to manage the peri-operative care of patients, recognise common complications and be able to deal with them or know to whom to refer. The trainee will be a safe and useful assistant in the operating room and be able to perform some simple procedures under minimal supervision and perform more complex procedures under direct supervision.

The MRCS examination has two parts: Part A (written paper) and Part B Objective Structured Clinical Examination (OSCE).

2.1 Part A (written paper)
Part A of the MRCS is a machine-marked, written examination using multiple-choice Single Best Answer and Extended Matching Items. It is a four hour examination consisting of two papers, each of two hours' duration, taken on the same day. The papers cover generic surgical sciences and applied knowledge, including the core knowledge required in all nine specialties as follows:

> Paper 1 - Applied Basic Science
> Paper 2 - Principles of Surgery-in-General

The marks for both papers are combined to give a total mark for Part A. To achieve a pass the candidate is required to demonstrate a minimum level of knowledge in each of the two papers in addition to achieving or exceeding the pass mark set for the combined total mark for Part A.

2.2 Part B (OSCE)
The Part B (OSCE) was introduced for first examination in September 2008. It integrates basic surgical scientific knowledge and its application to clinical surgery. The purpose of the OSCE is to build on the test of knowledge encompassed in the Part A examination and test how candidates integrate their knowledge and apply it in clinically appropriate contexts using a series of stations reflecting elements of day-to-day clinical practice.

3.   Changes to the Part B (OSCE) with effect from May 2010
The 2008/9 Annual Report announced the intention of the ICBSE to review the structure and content of the OSCE based on the evidence of the first three diets of the examination and that proposals for change would be submitted to the Postgraduate Medical Education and Training Board (PMETB) in autumn 2009. That review process began in May 2009.

It was based on the accumulated statistical evidence from over 850 candidates taking three diets of the MRCS Part B (OSCE): October 2008; February 2009; and, May 2009. ICBSE gathered qualitative feedback on all aspects of the examination from candidates, lay and medically-qualified examiners, assessors and invited observers including the Association of Surgeons in Training (ASiT). This information was analysed and evaluated by ICBSE's OSCE sub-group and by the Internal Quality Assurance Committee. Recommendations for modifications to the Part B (OSCE) emanating from that process were agreed by ICBSE in July 2009 and submitted to PMETB for approval as major changes to the April 2008 submission. Following a series of meetings with PMETB's Review Panels, at which amendments to the ISCP curriculum were considered alongside proposed changes to the MRCS, PMETB approval was granted early in 2010.

The changes that were agreed with effect from the May 2010 diet were as follows:
1. **The OSCE will comprise 18 examined stations (formerly 16).**
2. **The18 stations will be reconfigured into four Broad Content Areas (formerly five).**
3. **The knowledge, skills, competencies and professional characteristics to be assessed by the OSCE will be reconfigured into four domains (formerly six).**
4. **Domains will no longer represent pass/ fail criteria but will be used primarily for structuring the scenarios and mark sheets and for candidate feedback**
5. **Each station will be marked out of 20 (formerly 16).**
6. **All of the 18 stations will manned (formerly 12 of the 16).**
7. **The choice of specialty context stations will be reduced.**

The evidence base for these changes was as follows:

3.1 Number of stations
The MRCS Part B is more accurately to be described as a multi-station examination rather than an OSCE, in that it examines a wider range of disparate knowledge and skills than a more conventional OSCE. As a consequence, the use of Cronbach's alpha, an index of internal reliability, is likely to yield lower values for the MRCS (OSCE) than for examinations designed to sample a more cognate group of skills and attributes. Although the reliability of the MRCS Part B (OSCE) is acceptable based on comparator studies identified by PMETB[1], ICBSE acknowledges that it should strive to achieve a higher level on the index.

One way of improving reliability is to increase the number of stations and the length of the examination. Applying generalisability theory to the data available enabled ICBSE to predict the impact of increased examining time/number of stations on the achieved reliability coefficient (see Table 1 below). Using this information, together with the Utility Index promoted by ICBSE[2], it was decided that the optimum increase to the number of stations would be two with a concomitant increase in examining time of 20 minutes.

---

[1] based on the benchmarks provided in *Developing and maintaining an assessment system – a PMETB guide to Good practice,* January 2007,pg. 9, Box 1 Reliability as a function of time.
[2] based on the formula provided in *Developing and maintaining an assessment system –a PMETB guide to Good practice,* January 2007,pg. 7

Table 1: Application of Generalisability study to model the impact on reliability of increasing the number of OSCE stations

If the number of stations was…..	The reliability would be…
16	0.78
18	0.80
20	0.82
25	0.85
30	0.87
40	0.90

(Data from May 2009, N=432)

The increase in the number of stations, and the redistribution of stations between the Broad Content Areas, has enabled the ICBSE to add greater emphasis to the examination of the basic sciences – particularly anatomy and pathology – and in so doing has addressed the concerns raised in feedback on the examination by the majority of stakeholders.

3. 2 Broad Content Areas

The reliability of each of the BCAs was reviewed (see Table 2 below). The reliability of the Surgical Skills and Patient Safety (SSPS) BCA was an area of some concern.

Table 2: Reliability Coefficients for Broad Content Areas (BCAs)

Broad Content Area	Reliability coefficient
Anatomy and Surgical pathology (ASP)	0.59
Surgical Skills and Patient Safety (SSPS)	0.32
Communication Skills (CS)	0.52
Applied Surgical Skills and Critical Care (ASSCC)	0.49
Clinical (CLIN)	0.52

(Data from May 2009, N=432)

Since SSPS comprised only two stations it was judged that there was potential for a poor performance on one station to have disproportionate effects on a candidate's overall result. The continuing use of two station BCAs within the OSCE structure was therefore questioned and it was agreed to avoid the use of BCAs with two stations in the structure of the OSCE. Using the correlation pattern for the BCAs as a basis for decision-making (see Tables 3 and 4 below) it was decided to reconfigure the stations into four BCAs. These four new BCAs are as follows:-
Anatomy and surgical pathology
Applied surgical science and critical care
Communication skills
Clinical and procedural skills.

Table 3: Correlation pattern for five Broad Content Areas

	ASP	SSPS	CS	ASSCC
SSPS	0.24			
CS	0.22	0.43		
ASSCC	0.20	0.36	0.36	
CLIN	0.27	0.35	0.50	0.52

(Data from May 2009, N=432)

4

Table 4: Correlation pattern for Broad Content Areas dividing Clinical Skills into History Taking (HT) and Physical Examination (PE)

	ASP	SSPS	CS	ASSCC	HT
SSPS	0.24				
CS	0.22	0.43			
ASSCC	0.20	0.36	0.36		
HT	0.07	0.26	0.37	0.48	
PE	0.35	0.33	0.45	0.41	0.42

(Data from May 2009, N=432)

3.3/3.4 Domains

On the basis of evidence from the first three diets of the MRCS Part B, the construct validity of the domains was less compelling than that for the BCAs (see Table 5 below). This phenomenon had been identified following the analysis of the first examination diet and additional examiner training had been designed and delivered to underpin the conceptual basis of the domains and their operationalisation in the OSCE scenarios. The measurable impact of this additional examiner training on the construct validity of the domains was minimal. There was qualitative evidence from examiners to suggest that the conceptual distinctions between the six domains were, at times, too sophisticated and the allocation of marks to domains in the one minute between candidates was challenging. As a consequence, ICBSE agreed to remove them as pass/fail hurdles.

Table 5: Domains and construct validity

	October 2008	February 2009	May 2009
Adequately satisfied*	20 / 53	22 / 53	22 / 55

*Number of assessed domains satisfying criterion validity

There was, however, no evidence to suggest that domains were not a useful device for structuring questions and it was agreed that they should continue to be used for scenario design and for structuring the marking process. They will continue to form part of post examination candidate feedback. However, on the basis of the above evidence it was decided to reconfigure the six domains into four domains that more closely mapped to the syllabus. These four domains are as follows:

- **Clinical knowledge and its application**
- **Clinical and technical skill**
- **Communication**
- **Professionalism (including: Decision making, problem solving, situational awareness and judgement, organisation and planning, patient safety.)**

3.5 Marking
The reconfiguration of the OSCE assessment matrix has been accompanied by an increase in the number of marks available in each station from 16 to 20. This is to allow more refinement to the weightings given to the BCAs and domains in the marking system that continues to be based, fundamentally, on units of 4 i.e. 4: Pass, 3: Borderline pass, 2: Borderline fail, 1: Fail.

3.6 Use of Examiners

Evidence from the three diets of the examination demonstrated that unmanned stations showed statistically significant lower mean scores compared to manned stations (see Table 6 below). Although this was not, of itself, considered to be problematic, since every candidate was given a similar examination experience, the candidate feedback consistently raised the format and structure of these unmanned stations as an area of concern. In terms of utility, an examiner had been allocated to each of the unmanned stations for each of the first three diets to oversee the candidates and to mark their outputs. To convert these stations into stations manned by an examiner therefore presented no increase in human resources and, it was considered, would enable more flexibility in the knowledge, skills and attributes to be examined - particularly in the domains of communication and professionalism.

Table 6: Analysis of variance: manned/ unmanned stations

	Mean score ex 16	ANOVA result
Manned stations	12.5	$p = 0.0001$
Unmanned stations	9.5	

3.7 The choice of specialty context stations.

Based on the analysis of the first three OSCE diets, there was a statistically significant difference between the mean score of candidates in their chosen specialty context stations and their generic stations: the former was higher than the latter (see Table 7 below). This was, perhaps, to be expected. However, stakeholder feedback, particularly from ASiT, challenged the amount of choice offered to candidates on the basis that the OSCE was essentially an examination of the knowledge, skills and attributes to be expected of all trainees at that level. This is reinforced by the new syllabus. ICBSE agreed that candidate choice of specialty context stations would be reduced such that candidates could 'opt out' from one of the four specialty context stations. ICBSE considered that this was likely to improve the uniformity of candidate experience and contribute to an increase in the overall level of the internal reliability of the OSCE.

Table 7: Analysis of variance: generic/specialty context stations

	Mean score ex 16	ANOVA result
Generic stations	11.6	$p = 0.0001$
Specialty context stations	12.3	

These changes in the Part B (OSCE) examination were implemented in the May 2010 diet. For more information about the MRCS go to www.intercollegiatemrcs.org.uk.

4. The MRCS and the Intercollegiate Surgical Curriculum Programme (ISCP)

The MRCS examination is an integral part of the assessment system of the Intercollegiate Surgical Curriculum Programme (ISCP) http://www.iscp.ac.uk. Nine surgical specialties: cardiothoracic surgery; general surgery; neurosurgery; oral & maxillofacial surgery; otolaryngology; paediatric surgery; plastic surgery; urology; and trauma & orthopaedic surgery have collaborated through the ISCP in developing a competence-based curriculum which defines the attributes required of a successful surgeon. The web-based ISCP curriculum and its assessment system, including the MRCS, have been approved by PMETB (now GMC).

5. Writing the MRCS examination and standard setting

5.1 Part A written papers

Based on the ISCP curriculum, a syllabus blueprint for the Part A examination sets out a broad specification for the numbers of questions on each topic to be included in each paper of the examination. It is not possible to sample the entire syllabus within a single Part A paper but the blueprint and specification ensures that the common and important content is routinely covered and that the entire syllabus is sampled over time.

Questions are coded according to the area of the syllabus to which they relate and are held in a computerised item bank. During the year groups of question writers were commissioned to produce new questions according to the agreed specification and, following editing and specialist review, these questions were added to the item bank. For each diet of the examination questions are selected from the bank using the examination blueprint and are compiled into a paper by the MCQ question paper group of the ICBSE.

Questions are carefully planned from the outset to be at an appropriate level of difficulty. The standard for the paper is originally set using a modification of the Angoff procedure where a group of 'judges' estimates the performance of a notional 'just good enough to pass' candidate. In order to ensure that standards are set at an appropriate and realistic level the judges include practising surgeons, trainers, a trainee and a patient representative. A number of 'marker' questions taken from a previous examination are included in each Part A paper and are used to calibrate the standard and help to ensure that there is continuity between the standard of the examination over time.

Following each examination a standard setting meeting is held at which the performance of candidates on each question is scrutinised together with their performance on the test overall. A range of statistical measures is used to evaluate the reliability and facility of the examination and its individual questions. It is at this stage that candidate feedback on the examination is considered and taken into account when deciding whether or not to exclude a specific question from the overall examination outcome. Using the benchmark of the previously described Angoff exercise, the performance of candidates on the marker questions is reviewed together with other statistical data from the present and previous examinations to set the pass/ fail cut-off mark.

Candidates are given their Part A score and the score required to pass the examination, thus giving them an indication of how far short of, or above, the required standard they are.

5.2 Part B (OSCE)
Scenarios and questions for the OSCE stations are written by a team of Broad Content Area leads using detailed templates and a detailed writers' guide. Draft scenarios are scrutinised by a team of reviewers before being edited and submitted to a final review panel for approval. The scenarios are then grouped into examination 'circuits' so as to achieve the appropriate balance of content and challenge.

The same circuits are used in each of the Colleges on the same day. Circuits are changed each day. At the end of examination diet, the results of all candidates are combined and the pass/fail boundaries are agreed at a single standard setting meeting attended by representatives of each of the Colleges. Each standard setting meeting begins with an analysis of the level of discrimination and facility of each of the circuits and their constituent stations, including a review of candidate and examiner feedback, to ensure consistency and comparability of demand.

Each candidate's performance on each of the examined stations is assessed in two ways:
- a mark is awarded using a structured mark sheet containing assessment criteria for each content area and for each assessed domain;
- an holistic judgement using one of the categories: Pass; Borderline pass; Borderline fail; or Fail.

The following information is therefore available for each candidate:

- a total mark for each station;
- a category result for each station i.e. Pass; Borderline pass; Borderline fail; Fail;
- a total mark for the OSCE;
- a total mark for each domain;
- a total mark for each broad content area.

Using the above information, a variant of the *contrasting groups* method of standard setting [3] is used to determine the pass fail boundary for the OSCE as a whole as follows:

1. A *lower limiting mark* is calculated using the scores for all candidates awarded 'borderline fail'.

2. An *upper limiting mark* is calculated using the scores for all candidates awarded 'borderline pass'.

3. The Standard Error of Measurement (SEM) for the OSCE is calculated and added to the mid point between the upper and lower limiting marks. The resultant sum, rounded upwards, is normally taken as the pass/fail mark but a higher or lower mark may be chosen between the upper and lower limiting marks on the basis of examiner judgement and a review of available evidence.

4. To safeguard the interests of patients, and as a driver to learning, it is a requirement for passing the OSCE that in addition to achieving a pass mark in the OSCE overall, a minimum level of competence must be achieved in each Broad Content Area. For the diets held in October 2009 and February 2010 it was also a requirement that a minimum level of competence must be achieved in each domain. The minimum level of competence is an issue of examiner judgement based on a review of available evidence and is represented by a mark between the upper and lower limiting marks.

Each candidate is given detailed feedback showing their mark on each Broad Content Area, each domain and for the OSCE overall.

---

[3] Norcini, J. J. Setting Standards on Educational Tests, *Medical Education* 2003;37:464-469

6. Summary descriptive statistics: MRCS Part A

	Total number sat	Passing % (and number)	Failing % (and number)	Pass mark %	Measure of reliability*	Measurement error** %
Sept 2009	814	62.7 (510)	37.4 (304)	65.1	0.94	2.5
January 2010	632	59.3 (375)	40.7 (257)	64.3	0.93	2.7
April 2010	691	63.7 (440)	36.3 (251)	67.5	0.93	2.5

* An expression of the consistency and reproducibility (precision) of the examination. The measure used here is KR-20.

** Measurement error refers to the difference between the 'true' score and the score obtained in an assessment. Measurement error is present in all assessments but is minimised by good item design and test construction.

7. Summary descriptive statistics: MRCS Part B (OSCE)

	Total number sat	Passing % (and number)	Failing % (and number)	Pass mark %	Measure of reliability*	Measurement error** %
October 2009	479	50.1 (240)	49.9 (239)	66.0	0.80	4.4
February 2010	538	61.5 (331)	38.5 (207)	66.8	0.82	4.2
May 2010	500	52.2 (261)	47.8 (239)	67.8	0.84	3.9

* An expression of the consistency and reproducibility (precision) of the examination. The measure used here is Cronbach's alpha.

** Measurement error refers to the difference between the 'true' score and the score obtained in an assessment. Measurement error is present in all assessments but is minimised by good item design and test construction.

8. MRCS: Review and further development
The ICBSE continues to review and further develop the MRCS examination based on the evidence available and is committed to undertake a further review of the examination following the February 2011 diet.

All candidates and their trainers will be kept up-to-date about developments and changes to the MRCS via announcements on www.intercollegiatemrcs.org.uk.

9. The Diploma of Otolaryngology Head & Neck Surgery (DO-HNS)

The Diploma in Otolaryngology – Head and Neck Surgery (DO-HNS) was established as an intercollegiate examination in April 2008. Its purpose is to test the breadth of knowledge, the clinical and communication skills and the professional attributes considered appropriate by the Colleges for a doctor intending to undertake practice within an Otolaryngology department in a trainee position. It is also intended to provide a test for those who wish to practise within another medical specialty, but who have an interest in the aspects of where that specialty interacts with the field of otolaryngology. It is also relevant for General Practitioners wishing to offer a service in minor ENT surgery.

The Intercollegiate DO-HNS examination has two parts: Part 1 (written paper) and Part 2 Objective Structured Clinical Examination (OSCE).

> **Part 1 – Written Paper**
> This comprises Multiple True/False Questions and Extended Matching Questions in one paper to be completed in two hours.

> **Part 2 - OSCE**
> The Part 2 OSCE normally comprises approximately 25 bays normally of seven minutes' duration.

10. Standard Setting the DO-HNS examination

The DO-HNS standard setting procedure for the Part 1 written paper is very similar to that described above for the MRCS (see 5.1 above) and is based on an initial Angoff process, the use of marker questions and the scrutiny of individual items and statistics at a standard setting meeting.

The standard setting technique used in the OSCE to determine the pass mark is an Angoff process: all examiners determine a pass mark for each station based upon **the minimum level of competence expected of an ENT trainee at the end of his/her SHO/CT2/ST2 post before entry to higher surgical training or just at the start of higher surgical training.** Using this method, at least 12-15 examiners will ascribe a pass mark to each station, the marks are totalled and averaged and this then determines the region of the pass mark. The final pass mark is determined by inspection of the mark distribution around the Angoff pass mark.

11. Summary descriptive statistics
11.1 DO-HNS Part 1 Written

	Total number sat	Passing % (and number)	Failing % (and number)	Pass mark %	Measure of reliability*	Measurement error** %
August 2009	110	71.8 (79)	28.2 (31)	73.0	0.92	6.67
January 2010	103	71.8 (74)	28.2 (29)	75.3	0.91	6.43
March 2010	89	73.0 (65)	27.0 (24)	75.1	0.89	6.35

* An expression of the consistency and reproducibility (precision) of the examination. The measure used here is KR-20.

** Measurement error refers to the difference between the 'true' score and the score obtained in an assessment. Measurement error is present in all assessments but is minimised by good item design and test construction.

11.2 DO-HNS Part 2 (OSCE)

	Total number sat	Passing % (and number)	Failing % (and number)	Pass mark %	Measure of reliability*	Measurement error** %
October 2009	117	62.4 (73)	37.6 (44)	66.0	0.98	1.88
February 2010	83	41.0 (34)	59.0 (49)	68.5	0.78	3.12
May 2010	103	41.7 (43)	58.3 (60)	67.2	0.70	2.84

* An expression of the consistency and reproducibility (precision) of the examination. The measure used here is Cronbach's alpha.

** Measurement error refers to the difference between the 'true' score and the score obtained in an assessment. Measurement error is present in all assessments but is minimised by good item design and test construction.

12. Quality Assurance and the role of IQA
The quality of the MRCS and DO-HNS examinations is monitored by the ICBSE's Intercollegiate Internal Quality Assurance Committee (IQA). The IQA meets at least three times each year and receives, for each part of the examinations, the following information:

- overall pass rates and descriptive statistics for the latest diet and previous diets;
- pass/fail breakdown by candidates'
    - first language for the latest diet and previous diets;
    - gender for the latest diet and previous diets;
    - primary medical qualification for the latest diet and previous diets;
    - ethnicity for the latest diet and previous diets.

After each examination, every candidate is invited to complete an anonymous feedback questionnaire. Examiners are invited to complete similar questionnaires. The IQA receives and reviews the feedback from examiners and candidates and correlates them with the statistical information on the examination.

In its interpretation of the data on the examination, the IQA is advised and assisted by an independent Educational Consultant who analyses the information and writes a brief report on each part of the examination drawing any potential anomalies to the attention of the Committee for consideration and action.

The IQA Committee will refer matters which it considers to be in need of attention or further scrutiny to either the appropriate subgroups of ICBSE. It also makes regular reports and recommendations to the ICBSE which has overall responsibility for the MRCS and DO-HNS examinations.