

# A summative, Objective, Structured, Clinical Examination in ENT used to assess postgraduate doctors after one year of ENT training, as part of the Diploma of Otorhinolaryngology, Head and Neck Surgery

A B DRAKE-LEE, D SKINNER\*, M HAWTHORNE†, R CLARKE‡,  
AND THE EXAMINERS FOR THE DIPLOMA OF OTORHINOLARYNGOLOGY, HEAD AND NECK SURGERY

## Abstract

**Context:** ‘High stakes’ postgraduate medical examinations should conform to current educational standards. In the UK and Ireland, national assessments in surgery are devised and managed through the examination structure of the Royal Colleges of Surgeons. Their efforts are not reported in the medical education literature. In the current paper, we aim to clarify this process.

**Objectives:** To replace the clinical section of the Diploma of Otorhinolaryngology with an Objective, Structured, Clinical Examination, and to set the level of the assessment at one year of postgraduate training in the specialty.

**Methods:** After ‘blueprinting’ against the whole curriculum, an Objective, Structured, Clinical Examination comprising 25 stations was divided into six clinical stations and 19 other stations exploring written case histories, instruments, test results, written communication skills and interpretation skills. The pass mark was set using a modified borderline method and other methods, and statistical analysis of the results was performed.

**Results:** The results of nine examinations between May 2004 and May 2008 are presented. The pass mark varied between 68 and 82 per cent. Internal consistency was good, with a Cronbach’s  $\alpha$  value of 0.99 for all examinations and split-half statistics varying from 0.96 to 0.99. Different standard settings gave similar pass marks.

**Conclusions:** We have developed a summative, Objective, Structured, Clinical Examination for doctors training in otorhinolaryngology, reported herein. The objectives and standards of setting a high quality assessment were met.

**Key words:** Teaching; Educational Assessment; Otorhinolaryngology; Competency-Based Education; Great Britain

## Introduction

National post graduate medical assessments should fulfil the expectations of the profession, the educators and the public.<sup>1</sup> The Royal Colleges in the United Kingdom and Ireland have a long tradition of providing summative assessment for doctors within the British Isles. After the ‘Calman reforms’, a common surgical Member of the Royal Colleges of Surgeons examination replaced the previously used ENT Fellowship examination.<sup>2</sup> At the same time, the Diploma of Otorhinolaryngology, Head and Neck Surgery (DO-HNS) examination replaced the Diploma in Otolaryngology (DLO) examination.<sup>3</sup> The standard of the newer Diploma was set at the level of those wishing to pursue ENT after themed core training in surgery during one year of

basic training in the specialty. The Postgraduate Medical and Training Board approved the curriculum and the Diploma of Otorhinolaryngology, Head and Neck Surgery examination.

Professor Harden developed the Objective, Structured, Clinical Examination (OSCE) in Dundee in the mid- to late 1970s to assess clinical competence.<sup>4,5</sup> This style of examination had previously been used in North America to assess the medical knowledge and clinical skills of surgical residents (the US equivalent of senior house officers and specialist registrars), and also to set standards of practice.<sup>6,7</sup> These examinations have also been undertaken in a variety of other situations, including licensure, but had not previously been used in surgery in the UK.<sup>8–10</sup>

From the ENT Department, University Hospital NHS Trust, Birmingham, the \*ENT Department, Royal Shrewsbury Hospital, the †ENT Department, James Cook University Hospital, Middlesbrough, and the ‡ENT Department, Royal Liverpool Children’s Hospital, UK.

Accepted for publication: 17 March 2009. First published online 16 July 2009.

Comment has been made in the medical educational literature regarding the paucity of publications from hospital specialty colleges in the UK.<sup>11</sup> We would like to rectify this, and so present the development and findings of the Diploma of Otolaryngology, Head and Neck Surgery OSCEs conducted between May 2004 and October 2008.

## Materials and methods

### *Objective, Structured, Clinical Examination: construction and structure*

The Diploma of Otorhinolaryngology, Head and Neck Surgery Objective, Structured, Clinical Examination (OSCE) was developed out of a formative, 10-station, first year assessment introduced in the West Midlands in 1995, which was undertaken during the first specialist registrar year.<sup>12</sup> The Diploma in Otolaryngology clinical section was replaced by 25 stations: six clinical stations, plus 19 stations assessing written case histories, investigation results, examples of instruments and equipment, images of gross and microscopic pathology (to test recognition), and datasets (to test interpretation). Answers were scored by examiners who cross-checked against a structured answer sheet. Candidates were allowed seven minutes at each station.

### *'Marker' stations and content validity*

The stations chosen for a particular examination were selected from a matrix of possible options, to ensure appropriate coverage of the curriculum. Each OSCE contained between four and six stations replicated in sequential examinations, to enable an assessment of comparability. The mean scores for these 'marker' stations were calculated for both examinations and an overall ratio calculated. This enabled comparison of cohorts of candidates, and acted as an important checking mechanism for standard-setting.

### *Examiners*

The examining board was made up of 30 senior doctors working in the UK. All had received formal training in assessment from the education department of the Royal College of Surgeons of England, and all had received equality and diversity training. The panel included three radiologists, one pathologist, and one general practitioner who specialised in ENT.

### *Standard-setting and statistics*

We used a variety of different standard-setting exercises, initially using a modified borderline method for setting the pass mark. Subsequently, we used the Angoff method for two examinations, concurrently with a borderline method, and found little numerical difference. Therefore, we used only the Angoff method thereafter. We used a distribution method for one examination.<sup>13</sup>

All data were examined initially using Microsoft Excel software, and were then exported into Minitab version 13 software for basic descriptive statistics, and into Statistical Package for the Social

Sciences version 13 software to check for internal consistency. Cronbach's  $\alpha$  value and split-half statistics were used to calculate coherence.

The key to ensuring that the public has confidence in medical assessment is to make sure that examinations are measured against a defined standard. To this end, we used both borderline and Angoff methods are 'set'. The process involves the assessment and the imaginary candidate.

The modified borderline method uses examiners to classify each candidate's performance and to grade the results as 'pass', 'fail' or 'borderline'. The marks of all candidates considered borderline are averaged for each examiner taking part in the exercise. These marks are then averaged to set the pass mark. When we used this method, the marks of between six and 12 examiners were assessed.

Recently, we changed to the Angoff method. Here, the examination is assessed prior to its use with candidates. Each examiner determines the mark that they believe a borderline candidate would obtain. The average of examiners' scores is then used to set the pass mark.

Unfortunately, there are errors in every examination; however, there are a number of ways to determine an individual examinations consistency. Both Cronbach's  $\alpha$  value and split-half statistics assume a normal distribution of marks and determine whether this is so; these parameters are the Contemporary educational theory methods of determining internal consistency.<sup>13</sup> A perfect assessment will have a value of one. Any value over 0.8 is good, and 0.9 or more is excellent. Any of these statistical method measures candidate and examiner performance and examination construction. The UK Postgraduate Medical Education and Training Board requires a Correlation of more than 0.9 by any of the statistical methods looking at consistency.<sup>14</sup>

### *Criterion referencing*

Seventeen volunteer senior house officers who had completed between six months and two years of ENT work undertook the a pilot OSCE under normal conditions, and commented on the type of assessment and its appropriateness and fairness. These comments were used to construct the first OSCE and to assist with initial standard-setting.

### *External inspection*

The outline plans for the Diploma of Otorhinolaryngology, Head and Neck Surgery examination were approved by the Postgraduate Medical and Training Board, both as a stand-alone assessment and as part of the Member of the Royal Colleges of Surgery examination. The passing of the DOHNS is part of the criteria for entry into higher training.<sup>14</sup>

Ethical approval for above processes was not required.

## Results

### *Internal consistency*

The two different standard-setting methods used gave results which differed only slightly. Table I

shows that there was good internal consistency, with a Cronbach's  $\alpha$  value of 0.99 for all examinations. Split-half statistics showed a little more variation, with values ranging from 0.96 to 0.99.

*Setting the pass mark*

All examinations bar one had a very similar ratio between the marker station; Comparison of marker stations between exams had ratio ranged from 1:0.97 to 1:1.05, except for one examination for which the ratio was 1:1.12.

We used a modified borderline method to set the pass mark for all examinations except two. After extensive debate, we used a distribution method to set the pass mark for one examination, as the candidates had been very similar previously. We reintroduced formal standard-setting by the modified borderline method, as this was the examination with the highest ratio. We used the Angoff method in parallel for two examinations, and thereafter used it as our sole standard-setting exercise.

**Discussion**

While there is debate between educationalists and medical practitioners about the merits and validity of internal and external assessment, national medical assessment is professionally driven, and an examination pass is accepted by the lay members of various national bodies as a sign of competence (lay representatives of the Royal College of Surgeons of England, personal communication). External assessment will form part of the structure of ENT medical accreditation for the foreseeable future, as the Diploma of Otorhinolaryngology, Head and Neck Surgery examination has been passed as 'fit for purpose' by the Postgraduate Medical and Training Board. It is important that the public have faith in national medical assessment; this is why we decided to report the first Objective, Structured, Clinical Examination (OSCE) in surgery in the UK.

The content validity of ENT assessment was easy to ensure, as most ENT consultations are short and structured and may be easily divided into appropriate sections that can be mapped to the curriculum. Any assessment reflects what is known at the time it is performed, and is subject to error; likewise, OSCEs are also subject to error, but this may be addressed by careful planning.<sup>9,15-21</sup> Structured assessment stations remove many of the sources of such error, relating to both assessment design and examiner performance. Two further advantages of OSCEs are that their structure is similar to the real clinical environment and is very flexible.<sup>21</sup> While our OSCEs concentrated on clinical skills, the scope of such assessment may be extended to include surgical skills.<sup>9,22</sup>

In order to ensure public confidence, we have trialled a number of marking strategies and have decided on marking by single examiners, in line with others.<sup>23</sup> While patients may expect to be treated by the best clinician, standard-setting recognises that performance varies. The successful examination candidate should be competent to participate in clinical practice, which is why examination

TABLE I  
SUMMARY OF MARKS AND INTERNAL VALIDITY FOR THE SIX EXAMINATIONS

Parameter	May '04	Oct '04	May '05	Oct '05	May '06*	Oct '06	May '07	Oct '07	May '08
Candidates (n)	73	62	78	86	142	90	83	55	82
Marks									
- Total	368	438	429	455	522	499	519	511	526
- Range	213-307	241-339	234-338	222-342	245-393	218-371	265-384	267-371	277-433
- Mean	265	282	287	294	328	308	328	317	355
- Median	265	282	286.5	297	329	307.5	331	312	354
- SD	20	22	21	25	27	25	27	23	29
Cronbach's $\alpha$	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
Split-half	0.99	0.98	0.97	0.96	0.96	0.97	0.98	0.97	0.99
Mean - 1 SD	245	260	266	269	301	283	291	284	326
Pass (%)	66.6	59.4	62	59.1	57.7	56.7	56	56	62
Borderline	252	270	270	281	ND	297	315 <sup>†</sup>	306 <sup>†</sup>	348 <sup>‡</sup>
Passes (n (%))	60 (82)	42 (68)	65 (83)	59 (69)	125 (88)	60 (67)	58 (70)	37 (67)	46 (55)

Marks data are given to the nearest whole number, except for median values. \*This group had the best ratio of marker stations (1:1.12); the pass mark was raised accordingly. †Exam pass mark set using both borderline and Angoff methods; results were within three marks of each other. ‡Exam pass mark set using Angoff method alone. Oct = October; SD = standard deviation; ND = not done

methods are criterion-based rather than varying from examination to examination (as in the past). As explained in the Materials and Methods section, there are two approaches to setting the pass mark, one based on performance and the other on how the candidate should perform. We initially used a modified borderline method (i.e. performance-based) to set the pass mark for seven out of the nine examinations reported here, and we chose a candidate-centred approach in which candidates fell into one of three groups: pass, borderline or fail.<sup>23–26</sup> We then introduced the Angoff method (i.e. pre-assessment-based) for two examinations and found little difference. This latter is now our method of choice. The Medical Council of Canada requires a pass in key stations as well as achieving a general pass mark, but we did not use this approach.<sup>24</sup> The proportion of candidates passing OSCE varied between 67 and 88 per cent.

- **National medical assessments should fulfil the expectations of the profession, educators and the public**
- **This paper describes the process by which the clinical section of the Diploma of Otorhinolaryngology examination will be assessed using an Objective, Structured, Clinical Examination**
- **This process has resulted in a national assessment which is both ‘fit for purpose’ and subject to stringent controls, thus fulfilling the public’s expectation for a professional assessment**

Both norm based tests and most of the common analytic techniques (such as Cronbach’s  $\alpha$  value, split-half statistics and generalisability methods) assume a normal distribution. Our examination marks conformed to a normal distribution. Various statistical tools may measure how we achieved our goal of providing a robust assessment; the two used are described above. Standard-setting has undergone a change from normative-based to competency-based protocols. Internal validity is clearly indicated by two complementary methods, Cronbach’s  $\alpha$  value and split-half statistics; all our values were above 0.96 (Table I). Examiner consistency may be measured indirectly by the split-half statistic, with values of more than 0.9 being desirable and levels of 0.8 and above being acceptable; we obtained values of at least 0.96, indicating very good consistency.<sup>13</sup>

## Conclusion

Through the above process, we believe that we have set up a national ENT assessment that is fit for purpose, and have subjected it to stringent controls, thus fulfilling the public’s expectations for professional medical assessment.

## Acknowledgements

We would like to thank the Royal College of Surgeons of England for their support and encouragement in developing this assessment. No funding was received.

All authors contributed to the development of the assessment. A Drake-Lee and M Hawthorne helped develop the curriculum and syllabus. D Skinner developed the Diploma of Otorhinolaryngology, Head and Neck Surgery Objective, Structured, Clinical Examination from the West Midlands Objective, Structured, Clinical Examination, which he developed with A Drake-Lee. At the time of writing, D Skinner was chairman of the examiners. R Clarke was the immediate past chairman of the examiners, and helped develop the standard-setting exercises with the other three authors.

We would like to acknowledge the following Diploma of Otorhinolaryngology, Head and Neck Surgery examiners, all of whom were co-authors of this paper: D J Alderson, A K Bhattacharyya, N R Bleach, D A Bowdler, M W Bridger, V L Cumberworth, A J Drysdale, K L Evans, R M Evans (radiologist), E W Fisher, M Harries, T R Helliwell (pathologist), D G John, J E Kabala (radiologist), D G Lewis-Jones (radiologist), G A Morrison, D A Nunez, C H Raine, P J Robb, A K Robson, D Snow, P D Spragg, A Swift, P L Thomas (general practitioner), F Wilson and R P Youngs.

## References

- 1 PMETB principles of an assessment system for postgraduate medical training, 2004. In: [http://www.pmetb.org.uk/fileadmin/user/Standards\\_Requirements/PM\\_ETB\\_Scas\\_July2008\\_Final.pdf](http://www.pmetb.org.uk/fileadmin/user/Standards_Requirements/PM_ETB_Scas_July2008_Final.pdf) [23 June 2009]
- 2 Calman K. *Hospital Doctors: Training for the Future: the Report of the Working Group on Specialist Medical Training*. Heywood: Department of Health, 1993
- 3 Diploma in Otolaryngology – Head and Neck Surgery: Syllabus. In: [http://www.intercollegiatemrcs.org.uk/dohns/syllabus\\_html](http://www.intercollegiatemrcs.org.uk/dohns/syllabus_html) [23 June 2009]
- 4 Harden R, Gleeson F. Assessment of clinical competence using an Objective Structured Clinical Examination (OSCE). *Med Educ* 1979;**13**:1134–40
- 5 Harden R, Stevenson M, Downie W, Wilson G. Assessment of clinical competence using objective structured clinical examination. *BMJ* 1975;**1**:447–51
- 6 Schwartz RW, Witzke DB, Donnelly MB, Stratton T, Blue AV, Sloan DA. Assessing residents’ clinical performance: cumulative results of a four-year study with the Objective Structured Clinical Examination. *Surgery* 1998;**124**:307–12
- 7 Cohen R, Reznick RK, Taylor B, Provan J, Rothman A. Reliability and validity of the objective structured clinical examination in assessing surgical residents. *Am J Surg* 1990;**160**:302–5
- 8 MacRae H, Regehr G, Leadbetter W, Reznick R. A comprehensive examination for senior surgical resident. *Am J Surg* 2000;**179**:190–3
- 9 Patil NG, Saing H, Wong J. Role of OSCE in evaluation of practical skills. *Medical Teacher* 2003;**25**:271–2
- 10 Brailovsky C, Grand’Mason P, Lescop J. A large-scale multicentre objective structured clinical examination for licensure. *Acad Med* 1992;**67**:S37–9
- 11 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
- 12 Drake-Lee A, Skinner D, Reid A. An OSCE as part of the first-year SpR assessment in ENT. *Ann R Coll Surg* 2006;**88**(suppl):310–12
- 13 Ebel RL, Frisbie DA. *Essentials of Educational Measurement*, 5 edn. Englewood Cliffs: Prentice Hall, 1991;370

- 14 Modernising Medical Careers. In: <http://www.mmc.nhs.uk/pdf/ST3%20otolaryngology%20person%20specification%20final%20051207.pdf> [23 June 2009]
- 15 Frederiksen J, Collins A. A systems approach to educational testing. *Educational Researcher* 1989;**18**:27–32
- 16 Boulet J, Champlain A, McKinley D. Setting defensible performance standards in OSCEs and standardized patient examinations. *Medical Teacher* 2003;**25**:245–9
- 17 Norcini JJ. Setting standards on educational tests. *Med Educ* 2003;**37**:464–9
- 18 Kramer A, Muijtjens A, Jansen K, Dusman H, Tan L, van der Vleuten M. Comparison of a rational and empirical standard setting procedure for an OSCE. *Med Educ* 2003;**37**:132–9
- 19 Davis MH. OSCE: the Dundee experience. *Medical Teacher* 2003;**25**:255–61
- 20 Hodges B. Validity and the OSCE. *Medical Teacher* 2003;**25**:250–4
- 21 Adamo G. Simulated and standardized patients in OSCEs: achievements and challenges 1992–2003. *Medical Teacher* 2003;**25**:262–70
- 22 Patil NG, Cheng SW, Wong J. Surgical competence. *World J Surg* 2003;**27**:943–7
- 23 Wilkinson TJ, Newble DI, Frampton CM. Standard setting in an objective structured clinical examination: use of global ratings of borderline performance to determine the passing score. *Med Educ* 2001;**35**:1043–9
- 24 Smee S, Blackmore D. Setting standards for an objective structured clinical examination: the borderline method gains ground on Angoff. *Med Educ* 2001;**35**:1109–1010
- 25 Humphrey-Murto S, MacFadyen J. Standard setting: a comparison of case-author and modified borderline-group methods in a small scale OSCE. *Acad Med* 2002;**77**:729–32
- 26 McKinlay D, Boulet J, Hambleton R. A work-centered approach for setting passing scores on performance-based assessments. *Evaluation and the Health Professions* 2005;**28**:349–69

Address for correspondence:  
Mr Adrian Drake-Lee,  
Queen Elizabeth Hospital,  
Edgbaston,  
Birmingham B15 2TH, UK.

E-mail: [Adrakelee@aol.com](mailto:Adrakelee@aol.com)

---

Mr A Drake-Lee takes responsibility for the integrity  
of the content of the paper.  
Competing interests: None declared

---