

# Development and implementation of the specialty certificate examinations

John Mucklow

**ABSTRACT** – Following successful pilots in 2006, knowledge-based assessments for those engaged in specialty training have been developed and implemented in 11 medical specialties, by the Federation of Royal Colleges of Physicians in partnership with the specialist societies. Over 400 physicians have been involved in a project that has required recruitment and training of up to 25 question writers in each discipline, and the constitution of examining boards and standard setting advisory groups in each specialty. The assessments (now known as the specialty certificate examinations) are delivered by computer-based testing in centres throughout the UK and overseas. This paper analyses the outcome of 14 examination diets, sat by 948 candidates, of whom 72% were occupying UK numbered training posts. A total of 786 candidates sat the examination in the UK, 162 in overseas centres. Pass rates among UK trainees have generally exceeded 80%, with reliability coefficients well in excess of 0.8.

**KEY WORDS:** computer-based testing, postgraduate examinations, reliability, single-best-answer questions, specialty training

## Introduction

The successful development of performance assessments by the Royal College of Physicians (RCP) in 2004, to support the new curricula for higher medical training, was followed by a project by the Joint Committee on Higher Medical Training (JCHMT) to introduce and pilot a knowledge-based assessment for specialist registrars, to complement workplace-based assessments of competence. Over a two-year period, assessments were developed in four specialties (cardiology, dermatology, geriatric medicine and neurology) by JCHMT in collaboration with the relevant specialist societies, and the pilots took place in 2006. The question format, procedures and standards adopted were modelled on those in use by MRCP(UK). Three specialties each set a single paper of 100 questions, the fourth (cardiology) a single paper of 50 questions. All endeavoured to recruit volunteer candidates from every year of the training programme and some who had completed their training. The outcome confirmed the feasibility of developing such assessments and established the requirements for their reliability.<sup>1</sup> This led to the proposal

that similar assessments be developed in a total of 13 medical specialties (Table 1). Responsibility for this passed to MRCP(UK) Central Office in view of its long experience of implementing assessments of this type and work began at the beginning of 2007.

## The assessment

Based on the conclusions of the pilots, a number of decisions were made about the design and delivery of each specialty's assessment. Each would comprise two papers of 100 single-best-answer (one-from-five) questions, which trainees would be advised to sit during the penultimate year of their specialty training. The assessment would be computer based rather than a paper test, to allow trainees to sit at a location reasonably close to home, to facilitate delivery of the assessment overseas and to streamline the marking process. Though initially referred to as knowledge-based assessments, since 2009 the assessments have been known as the specialty certificate examinations (SCEs). They are now a necessary requirement for the award of a certificate of completion of training (CCT) for all trainees who enrolled in training in one of the relevant specialties after the beginning of August 2007 (with the exception of acute medicine, for which the equivalent date is August 2009).

**Table 1. Specialties involved in developing specialty certificate examinations.**

Acute medicine
Cardiology*
Clinical pharmacology and therapeutics (until 2008)
Dermatology
Endocrinology and diabetes
Gastroenterology
Geriatric medicine
Infectious diseases
Medical oncology
Neurology
Nephrology
Palliative medicine (from 2010)
Respiratory medicine
Rheumatology

\*The British Cardiovascular Society have developed an independent specialist knowledge-based assessment.

John Mucklow, associate medical director, MRCP(UK) Central Office, Royal College of Physicians, London

## Coordination, recruitment and training

Specialist societies (Table 2) were approached and asked to identify a representative to coordinate recruitment from among their members of question writers and potential members of examining boards and standard setting advisory groups. Up to 25 question writers in each specialty were recruited and invited to a series of one-day training workshops. These explained the purpose and proposed delivery of the assessments, and the processes of question development and analysis, and provided expert guidance on the drafting of single-best-answer questions.

## Question writing, processing, review and storage

The process adopted to review drafted questions was modelled on that used in the preparation of single-best-answer questions for the MRCP(UK) written examinations since 1999. Members of specialist groups meet face to face for critical review of the content and design of each other's drafts over a two-day period, during which small groups (usually of six to eight individuals) can each review and approve around 100 new questions. Although expensive, this model has proved itself rigorous and reliable for the acquisition and maintenance of a large bank of questions for use by examining boards.

For the SCEs, in advance of each peer-review meeting, each question writer was asked to draft 15 new questions and submit these to a team of non-medical editors to allow preliminary vetting of question format and style. Each specialty group met every six months for the first 18 months and subsequently once a year. Writers found that their skill in drafting questions and avoiding the pitfalls grew with each meeting following the training day, and by the third peer-review encounter most were both confident and competent. Question accrual has varied among specialties, chiefly owing to the number of available

writers, but several specialties that began to produce questions in 2007 have now amassed more than 1,000 questions and the total bank now contains well over 10,000. These have been lodged in the electronic bank used for storage of MRCP(UK) questions pending the availability of a larger facility in 2011.

To facilitate storage and retrieval, each question has been coded according to the heading in the curriculum to which it relates, the clinical condition addressed in its content and the domains of knowledge tested.

## Examining boards

Each specialty was asked to identify 10 individuals to serve as examining board members, of whom not more than half should be current question writers; two should have been members of the relevant specialist advisory committee, and one of these should be a current member. They were also asked to identify a chair and secretary to oversee the board's activities.

Examining boards were convened as soon as each specialty had amassed 350–400 questions. Members were sent around 250 questions to review about three weeks before the meeting and then met for two days to consider their suitability for inclusion and to draft feedback to question-writing groups in respect of unsuitable questions. Each board's officers prepared a blueprint for approval by the board, setting out the distribution of questions in relation to topics in the specialty curriculum, and this governed the design of the papers. Given the curriculum-specific coding of each question, each diet of the examination can now be mapped to the specialty's curriculum. Examining boards now meet annually about eight months before the next diet.

## Standard setting advisory groups

Each specialty was asked to identify eight individuals to serve as members of a standard setting group (SSG); not more than half should be members of the examining board (to include the chair and secretary of the board). Ideally, the chair of the group should be an academic with experience of standard setting. SSG meetings took place two to three months before the diet. Members were sent around 220 questions to review about three weeks before the meeting and then met for two days to agree the level of difficulty of each question. At the first meeting of each SSG, a training session took place, followed by an opportunity to practise using questions not for use in the coming examination. The procedure involved using the modified Angoff method, which requires each member to consider independently the percentage of just-passing (borderline) candidates that ought to answer each question correctly. At the meeting, these independent judgements are discussed and reviewed, and members revise their assessments to approach a consensus. The mean of the mean values for all the 200 questions eventually selected for the examination becomes the criterion-referenced pass mark. The SSG meeting affords a further opportunity to consider each question's suitability for inclusion, and often involves substitution of 'spares' for those rejected.

**Table 2. Specialist societies collaborating in the development of specialist certificate examinations.**

Society for Acute Medicine*
British Pharmacological Society (until 2008)
British Association of Dermatologists*
Society for Endocrinology*
Diabetes UK*
Association of British Clinical Diabetologists*
British Society of Gastroenterology*
British Geriatrics Society*
British Infection Association
Association of Cancer Physicians
Association of British Neurologists
The Renal Association*
Association for Palliative Medicine (from 2010)
British Thoracic Society*
British Society for Rheumatology

\*Financial partners.

It is worth noting that each question selected for final inclusion in an SCE has been reviewed by as many as 20–30 specialists and discussed on three separate occasions (question-writing peer-review meeting, examining board meeting and SSG meeting). Every attempt is made at these meetings to achieve a consensus regarding the readability and clarity of each question, the essential nature of all its component parts, the correctness of the answer key and the relative incorrectness of the alternative options (the distractors). Without consensus, a question cannot be approved for use. Thus, the likelihood that a question whose answer remains debatable will appear in the examination is very small indeed.

## Eligibility and postnominals

Following discussions with the specialist societies, it was agreed that those wishing to enrol as candidates for the SCE should have completed the MRCP(UK) diploma or be occupying a UK numbered training post, or both. From 2011, it will no longer be necessary to have completed the MRCP(UK) diploma before enrolling as a candidate in an SCE (with the exception of dermatology and geriatric medicine). Successful candidates are awarded a certificate. Those who subsequently complete all the assessments of competence and fulfil the requirements for a CCT, or whose applications for a Certificate of Eligibility to the Specialist Register (CESR) prove successful, may apply to the Joint Royal Colleges of Physicians' Training Board (JRCPTB) for permission to use the postnominal, MRCP(UK) (name of specialty).

## Examination diets 2008–10

Delivery of the computer-based examinations has, in general, been a success. Pearson VUE, the commercial provider, has

access to centres throughout the UK and abroad where SCE candidates may sit the two papers. Three hours are allowed for each paper, with an interval between. The examination is sat in all centres and time zones simultaneously and strict security is observed within the centres by the invigilators.

The original intention was to deliver two diets per year in each specialty but this did not prove possible in view of the rate of question generation, and the professional time and expense involved. The outcome of diets held before September 2010 is shown in Table 3. In all, 948 candidates have sat an SCE, of whom 683 (72%) were UK trainees (occupying UK numbered training posts), all of whom sat the examination in the UK. Of the remainder, 39% sat the examination in the UK, while the others attended centres overseas, chiefly in Dubai (24%), Saudi Arabia (20%) or Egypt (8%).

The pass rates during 2008 and 2009 reflected the small cohorts of candidates, unfamiliarity with the content of the examinations and the standard required, and a lack of adequate preparation. Overall numbers of UK trainees enrolling have increased during 2010, and now include a significant number of candidates whose training began after August 2007. Pass rates among UK trainees in most specialties now exceed 80%.

Overseas candidates who are not UK trainees comprise chiefly graduates of medical schools in the Indian subcontinent (42%) or Africa (31%), who markedly outnumber those graduating from schools in the UK, Eire or Australia (12%), or the Middle East (10%). These candidates are appreciably older than UK trainees. Whereas two-thirds of UK trainees had graduated in 2000 or later, four-fifths of overseas candidates had graduated before 2000, and one-third before 1990.

The final column of Table 3 presents the reliability coefficient (Cronbach's  $\alpha$ ) for the diets to date. This assesses the likely

**Table 3. Outcome of examination diets to August 2010.**

Specialty	Date of exam	Number of candidates (UK trainees)	Pass mark (%)	Overall pass rate (%)	UK trainee pass rate(%)	Non-UK trainee pass rate(%)	Reliability
Gastroenterology 1	Jun 08	8 (6)	66.76	62.5	83.3	N/A	0.84
Geriatric medicine 1	Mar 09	15 (12)	59.41	100.0	100.0	100.0	0.48
Nephrology 1	Mar 09	33 (16)	66.88	57.6	81.3	35.3	0.86
Respiratory medicine 1	Apr 09	25 (14)	64.27	60.0	85.7	27.3	0.85
Neurology 1	May 09	25 (16)	56.46	80.0	87.5	66.7	0.89
Endocrinology/ diabetes 1	May 09	39 (14)	65.95	38.5	64.0	24.0	0.89
Dermatology 1	Sep 09	39 (30)	67.58	92.3	97.1	66.7	0.88
Infectious diseases 1	Sep 09	6 (0)	63.71	0.0	N/A	0.0	0.94
Gastroenterology 2	Nov 09	105 (78)	64.91	61.0	61.5	59.3	0.81
Geriatric medicine 2	Mar 10	160 (154)	58.00	83.1	83.1	83.3	0.74
Nephrology 2	Mar 10	97 (65)	63.00	60.8	80.0	21.9	0.83
Respiratory medicine 2	May 10	147 (125)	56.78	68.0	74.4	31.8	0.81
Neurology 2	May 10	75 (55)	53.03	73.3	85.5	40.0	0.91
Endocrinology/ diabetes 2	Jun 10	174 (98)	60.00	69.5	86.7	34.2	0.89

consistency of the pass rate if the same cohort of candidates were to re-sit the same papers soon after the original attempt. The formula used to calculate reliability depends on the variation in candidate performance and is strongly influenced by the range of candidate ability within each cohort. The pilot knowledge-based assessments, despite their structured cohorts, yielded reliability values no greater than 0.81. Reliability is also influenced by the number of questions in the examination; analysis of pilots concluded that a single paper of 100 questions was insufficient and anticipated the need for 200 questions if higher values were to be attained consistently.<sup>1</sup> Reliability values in the SCEs have generally exceeded 0.8. Ideally, the reliability figure should exceed 0.9 in high-stakes examinations, but this figure is difficult to achieve at this level of professional attainment, even in large cohorts, as candidates have already proved their ability to pass MRCP(UK) in its entirety and have a great deal of knowledge in common. Partly for this reason, reliability may not be the optimal measure of quality in postgraduate medical assessments, especially when these involve small cohorts, and some have argued that the standard error of measurement should be preferred.<sup>2</sup>

## Conclusion

SCEs have been developed and implemented in 11 medical specialties over the course of four years. After the initial shock of a further knowledge-based assessment, trainees are now accepting the examinations as a necessary component of their training programmes. The opportunity to test their specialist knowledge

in an examination developed to the same standards as MRCP(UK) should prove increasingly attractive to overseas trainees. Unfortunately, it has not been feasible to use this model of assessment for medical specialties with fewer than 30 trainees per year, and the smallest specialties are currently exploring alternative assessment options.

## Acknowledgements

Over 400 physicians from throughout the UK have taken part in the development of the SCEs and the success of the venture is largely the result of their diligence and devotion. I am especially grateful to the lead specialists who have given so much time to steering the project to a successful conclusion, and to the officers of the examining boards and standard setting advisory groups.

## References

- 1 Booth J. Knowledge-based assessment pilot project. *Clin Med* 2007;7:9–11.
- 2 Tighe J, McManus IC, Dewhurst NG, Chis L, Mucklow J. The standard error of measurement is a more appropriate measure of quality for postgraduate medical assessments than is reliability: an analysis of MRCP(UK) examinations. *BMC Medical Education* 2010;10:40.

**Address for correspondence: Dr J Mucklow,  
MRCP(UK) Central Office, 11 St Andrews Place,  
Regent's Park, London NW1 4LE.  
Email: [john.mucklow@mrcpuk.org](mailto:john.mucklow@mrcpuk.org)**

## Working party report

### Hearing and balance disorders

#### Achieving excellence in diagnosis and management

This report, by a multidisciplinary working party, proposes a coherent networked model of care, with new working patterns. It discusses future manpower needs, for the care of both the adult and paediatric populations with hearing and balance needs, and outlines the improvement in training across all levels and disciplines of medicine and related healthcare professions, advocating the integration of training and assessment of competencies.

The recommendations aim to produce and appropriately skilled workforce providing rapid access to care within the community in a cost-effective manner, basing the care on the restructured audiology and medical workforce, and ensuring that multiple referrals, inappropriate and expensive investigations and hearing and balance-related absence from work are kept to a minimum. ■



**Royal College  
of Physicians**

**Published: January 2008 ISBN: 978 1 86016 300 5**  
**Price: £20.00 UK, £22.00 overseas (inc post and packing)**  
**10% discount for fellows and members (quote ref *Clinical Medicine*)**