

Knowledge-based assessment pilot project

Joe Booth

In recent years there has been a great deal of work within the Joint Committee on Higher Medical Training (JCHMT), the Royal Colleges and medical education in general to investigate and implement new methods of assessment. A successful pilot study of three methods of performance assessment for specialist registrars (SpRs) was completed by the Royal College of Physicians (RCP) in 2004 and these methods are now being implemented. A JCHMT project, to pilot the use of knowledge-based assessment for SpRs, was run between 2004 and 2006, in partnership with specialist societies. The philosophy behind this second project lay in the expectations of patients, the public and the Government that specialists dealing with them have the appropriate knowledge of the specialty in which they practise as independent practitioners. The objectives of the project were:

- to consider the reliability and validity of the proposed assessment method
- to understand the practical implications of introducing knowledge-based assessment.

Cardiology, dermatology, geriatric medicine and neurology introduced pilot assessments. A project board was established for each specialty, with representation from both the specialist advisory committee (SAC) and the relevant specialist society (British Cardiovascular Society, British Association of Dermatologists, British Geriatrics Society, and Association of British Neurologists).

Principles

Five project principles were initially established and subsequently amended:

1 *Assessment to be formative with a summative element*

The original idea was that trainees would have an online bank of questions to be used formatively to guide their learning. At some stage they would take a summative test in a secure, invigilated setting and it quickly emerged that summative assessment was the key.

2 *Use 'best-of-five' format, multiple choice questions*

Best-of-five questions consist of a descriptive stem, a specific lead-in question and five options from which

examinees must choose one best answer. This format has replaced true/false questions in MRCP(UK) parts 1 and 2 and was therefore an obvious choice for this project. The Colleges have built up considerable expertise in writing this type of question.

3 *Web-based assessment*

The pilot assessments were initially a paper-based examination although the pilot specialties are all supportive of computerised delivery in the future.

4 *Core elements of specialty curricula*

Question setters were asked to restrict themselves to core components of the Higher Medical Training curriculum for the specialty. Participants were told that questions would be based on these curricula.

5 *Timing of assessment*

Specialties in the pilot anticipated that SpRs would take assessments in year 2 or 3, with the option of retakes as necessary in order to pass, before penultimate year assessment.

Question/exam setting process

The question setting and the pilot exam paper setting were modelled on the MRCP(UK) process, though with some differences between specialties. Standards for the type, style, format, wording etc of questions were created based on the standards already in place for MRCP(UK). These were supported by templates, examples and tips. Question setters were identified by a combination of personal approaches from board members and requests for volunteers. Writers were invited to question setting days at the RCP and were asked to contribute questions in advance. These were then reviewed and edited by small groups on the day. A draft paper was selected, reviewed and edited further by a group of board members and question setters and finally proofread. Cardiology set 50 questions while the other specialties used 100. Cardiology, dermatology and neurology set a notional pass mark, based on the Anghoff method, in advance of participants taking the assessment. A standard setting group of about six people was established. Each standard setter rated each question according to what percentage of 'just-passing' candidates should get the question right. A standard setting meeting was then held, each question was briefly

Joe Booth BSc,
Education Projects
Manager, Royal
College of
Physicians, London

Clin Med
2007;7:9–11

Table 1. Number of participants per specialty.

Specialty	SpR					Consultants	Other	Total	SpR participation* (%)
	Year 1	Year 2	Year 3	Year 4	Year 5				
Cardiology	61	83	60	51	25	12	11	303	52
Dermatology	46	53	53	29	NA	33	6	220	83
Geriatrics	83	79	90	78	72	37	9	448	77
Neurology	47	33	43	39	34	30	0	226	97

*As a percentage of the number of SpRs registered with JCHMT in July 2006. JCHMT = Joint Committee on Higher Medical Training; NA = not applicable; SpR = specialist registrar.

discussed and raters were asked to support their views. After the discussion, each group member was then asked to re-rate the question. These ratings were recorded, the mean calculated for each question, and the mean of these figures taken to give the pass mark. One of the problems with standard setting was establishing a common idea of a ‘just-passing’ candidate, and the expected level of knowledge. In neurology this was explicitly agreed to be the knowledge expected of a SpR upon successful completion of training. For dermatology it was based on the core dermatology knowledge expected of a SpR at the start of year 3 training.

Pilot operation

Pilot assessments were held locally in all UK regions. The pilot assessments took place on different days for the four different specialties. Assessments took place in exam conditions with invigilators present to prevent cheating, copying or any collaboration, and to ensure that question papers were not copied or taken away. All SpRs were strongly encouraged to take part along with consultant volunteers to provide a range of data. We guaranteed anonymity to participants by allocating random identifying numbers. Participants would then use their number to identify their own result from details published on the JCHMT website. Questionnaire forms were used to gather demographic data about participants and comments. After marking, participants were allocated to deciles according to their performance within their year/post group and also within the total population of participants. Results for each specialty were published on

the JCHMT website about one month after each pilot had taken place.

Results

The numbers of participants for each specialty (Table 1) and the mean scores of participants grouped by type of post and year of training (Table 2) were recorded. Table 3 shows the percentage of participants by specialty and group who would have passed the exam had the standard setting pass mark been applied. No pass mark for geriatric medicine was set by formal standard setting but a notional pass mark of 50% was applied after the paper was taken. The pass mark for cardiology was clearly too high, probably because the standard setting process was not followed rigorously and was combined with question editing and selection processes. Both dermatology and neurology were happy that their pass marks performed well in identifying a minority of trainees with possible deficiencies of knowledge at the year 2/3 stage.

Reliability

The reliability of a multiple-choice examination can be measured using Cronbach’s coefficient alpha. This is a measure of internal consistency which indicates how much of the gathered information is ‘true’ information and how much is error. It is a number between 0 and 1, with higher values indicating a more reliable examination. High-stakes examinations are usually expected to have an alpha value exceeding 0.80, and preferably higher than 0.90.

Table 2. Mean percentage scores.

Specialty	SpR					Consultants	Overall
	Year 1	Year 2	Year 3	Year 4	Year 5		
Cardiology	68.03	69.04	70.07	72.74	71.12	74.67	70.01
Dermatology	63.37	69.36	73.11	72.48	NA	77.91	70.74
Geriatrics	56.77	58.11	58.59	58.65	59.90	62.62	58.56
Neurology	59.00	64.33	64.49	68.89	70.50	71.40	65.92

NA = not applicable; SpR = specialist registrar.

Table 3. Percentage of participants passing the exam.

Specialty	Pass mark (%)	SpR					Consultant	Overall
		Year 1	Year 2	Year 3	Year 4	Year 5		
Cardiology	83	0.0	2.4	8.3	7.8	7.7	8.0	4.8
Dermatology	60	63.0	88.7	98.1	96.7	NA	97.0	87.3
Geriatrics	50	83.1	88.8	91.1	96.1	91.7	97.3	87.5
Neurology	57	59.6	84.8	83.7	92.1	100.0	96.7	84.5

NA = not applicable; SpR = specialist registrar.

Specialty	number of questions		
	50	100	200
Cardiology	0.55	0.71	0.83
Dermatology		0.81	0.90
Geriatrics		0.67	0.81
Neurology		0.78	0.88

Reliability is affected by three major factors – the number of questions, the quality of the questions and the spread of ability of candidates. If all these factors hold true then the reliability of a longer exam can be estimated using the Spearman-Brown formula. Table 4 shows the calculated reliability of the pilots (numbers in bold) and extrapolations of what the reliability of a longer examination with similar characteristics might be.

Conclusions

The project demonstrated that setting high quality specialty examinations using best-of-five questions is a feasible activity for the Federation of Royal Colleges of Physicians working in partnership with specialist societies, and that it is possible to achieve acceptable levels of reliability using this method.

The specialties who took part in the project have a clear

understanding of the complexity and scale of setting a specialty examination. The setting and operation of high quality specialty examinations requires considerable effort, experience and expertise. It is appropriate that the existing MRCP(UK) organisation should take on this responsibility. The existing project boards and question setting groups are well placed to take this work forward for these specialties, though in some cases they may need reorganising to boost expertise and ensure suitable representation of stakeholders. Specialist examinations should ideally be based on 200 questions to maximise the reliability. This will, however, present a challenge to question setting groups and may not be achievable immediately.

Acknowledgements

This project was entirely funded by a grant from the Department of Health.

Further details

The full project report, including more details of results for each specialty and more discussion of issues arising from the project, can be accessed from the JCHMT website:

www.jchmt.org.uk/assessment/knowledgeAssessment.asp