J R Coll Physicians Edinb 2012; 42:101–5
doi:10.4997/JRCPE.2012.202
© 2012 Royal College of Physicians of Edinburgh

Improving the efficiency of selection to Core Medical Training: a study of the use of multiple assessment stations

[1]JM Atkinson, [2]E Tullo, [3]H Mitchison, [4]MS Pearce, [5]N Kumar
[1]ST6 Acute Medicine, North Tees General Hospital, Stockton-on-Tees, UK; [2]Teaching Fellow, UK NIHR Biomedical Research Centre in Ageing, Institute for Ageing and Health, Newcastle University, UK; [3]Consultant Physician, Sunderland Royal Hospital and former Head of School of Medicine, Northern Deanery; [4]Reader in Lifecourse Epidemiology, Institute of Health and Society, Newcastle University, UK; [5]Consultant Physician and Honorary Senior Clinical Lecturer, University Hospital of North Durham and Foundation School Director, Northern Deanery

ABSTRACT

Purpose: To compare three separate assessment stations used for selection to Core Medical Training (CMT) and to determine the effect of reducing the number from three to two.
Methods: Quantitative analysis of candidates' assessment station scores, financial analysis of costs of the selection process and quantitative and qualitative surveys of candidates and assessors.
Results: The assessment stations used for selection to CMT were reliable and valid for assessing suitability for employment as a CMT trainee. There was no significant difference in candidate ranking if only two assessment stations were used rather than three, i.e. there was no change in the likelihood of receiving a job offer. All of the assessment stations were perceived to have face validity by candidates and assessors. The efficiency of the selection process could be improved without loss of quality if two stations were used rather than three.
Conclusions: Using two assessment stations rather than three would appear to improve the efficiency and maintain the quality of the CMT selection process while reducing costs.

KEYWORDS Selection, trainees, interviews, physicians, Modernising Medical Careers

DECLARATION OF INTERESTS Dr Kumar has received payment from Roche for development of educational presentations and owns shares in GSK.

Correspondence to ***E Tullo***
Institute for Ageing and Health, Campus for Ageing and Vitality, Newcastle University NE4 5PL, UK

tel. ***+44 (0)191 248 1300***
e-mail ***ellen.tullo@ncl.ac.uk***

INTRODUCTION

'The selection of the doctors of tomorrow is a subject of constant interest because it raises questions about ensuring equity, predicting human behaviour, and defining the characteristics of a good doctor'.[1] In response to *Unfinished business*[2] the Department of Health and the Postgraduate Medical Education and Training Board (PMETB) developed a series of reforms, known as Modernising Medical Careers (MMC), which were launched in 2005. These reforms introduced a series of radical changes, with the aim of improving postgraduate medical education and training. Although the implementation of MMC was widely criticised, the overhaul stimulated progressive standardisation of selection for postgraduate specialties including Core Medical Training (CMT).[3–4] The CMT selection process has evolved since then. Initially uncoordinated and controlled locally by Deaneries, the process later came under the direction of the Royal College of Physicians of London (RCPL) with candidates being selected locally. In 2010 it became a nationally coordinated recruitment process, administered by the Joint Royal Colleges of Physicians Training Board (JRCPTB).[5]

Selection for medical training necessitates a system of ranking candidates in order to offer a set number of posts to the strongest candidates. Professional examinations aim to assess a minimum level of existing competence and it is therefore possible for all candidates to 'pass'. In contrast, when selecting for training posts, desirable candidate criteria might include potential ability, and clearly not all candidates will meet the requirements to be offered a post.[6] Evaluating the process of ranking candidates for selection to medical posts is notoriously complex and involves assessment of reliability, validity, fairness and cost. There is little long-term information available to establish whether we are selecting the right candidates. Validation of any selection method is likely to involve more than one study design. It is also recommended that candidate reaction be recorded in order to determine the perceived fairness of the process among participants.[6]

AIMS

In 2007 and 2008 the interview process in the Northern Deanery (Cumbria and the North East of England) consisted of three locally-devised assessment stations: a structured interview, a case-based discussion and a

communication station, all described elsewhere.[7] In 2009 the selection process evolved in accordance with the national scheme led by the RCPL.[8] Despite the inherent difficulties involved in the evaluation of a selection process, the changes implemented in accordance with the RCPL scheme prompted us to review our selection process. Our aim was to use a multi-method approach to establish the efficiency of selection performed in 2009, to assess whether a reduction in the number of stations changed the ranking of the candidates and to examine the opinions of applicants and assessors on the process.

METHOD

Description of the selection process in 2009

In 2009, a national scheme (allowing local shortlisting followed by local interview) led by the RCPL was introduced.[7] Three stations, each lasting for ten minutes had two assessors who each contributed an independent mark, resulting in two marks per task. There was a structured interview including a review of candidate portfolio, a case-based discussion (involving assessment of communication skills) and a third station consisting of a postulated scenario (with an actor) focused on a communication or ethical issue, including the topic of professionalism. Each candidate was required to complete all three assessment stations and the marking scheme for each station was structured. Participants were given clear instructions, there were indicators where marks should be given and positive and negative discriminators were outlined. A different pair of assessors was present at each station.

The scores for each station were anonymised and collated on a spreadsheet by a representative of the Northern Deanery. Each assessor gave a score from one (lowest) to five (highest) in two different domains. The scores from each of the two assessors from each of the three stations were totalled to give a final score by which the candidates were ranked.

We compared each score from the three stations for each candidate to assess their impact on the overall ranking and therefore success in the recruitment process. In 2009 the final candidate ranking also included a score based on the quality of the application form and each station was also given a weighting by the RCPL. In this paper we consider only the overall ranking based on the results from the station assessments before these additions were made.

Evaluation of selection process

Statistical analysis included the use of the Friedman test, a one-way analysis of variance (ANOVA), Wilcoxon signed-rank tests and Pearson and Spearman correlations. Statistical significance was considered to be $p<0.05$.

Evaluation of opinion of selection process

Assessor survey

A semi-structured questionnaire was sent to all assessors after all of the selection days using an online survey instrument.[9] The majority of questions used a Likert scale with closed questions to assess perception of how discriminatory the stations were and which station the assessors considered to be the most discriminatory. There was also an open section in the questionnaire to allow additional comments.

Candidate survey

All 2009 CMT candidates were also asked to complete a questionnaire with closed questions regarding their opinions as to the most discriminatory station, as well as questions regarding logistical aspects of the day.

RESULTS

Scores

In 2009, 183 candidates for CMT year one were interviewed using three assesment stations (Table 1). There was a consistent standard deviation of approximately one across all the stations.

There were statistically significantly high correlations between the scores of the two independent assessors (Pearson's correlation coefficient [r], $p<0.0001$). Scores were therefore amalgamated and a mean score was used for the ranking of the candidates.

Table 2 illustrates the correlations of the stations with each other, when comparing the candidate's actual scores. Candidates received similar scores across the stations, indicating consistent performance measurement.

The correlations between each station and the overall ranking (Table 3) are negative because as the score of the candidate within a station increases, so the ranking number decreases (one is the highest rank). The stations that correlated best with the rank were the ethical scenario and the professionalism station.

All stations correlated significantly well with each other in terms of the rank of the candidate, and no one station altered the candidate's likelihood to receive a job offer.

Costs

The assessment of the 183 CMT doctors in 2009 required a total of 147 consultant sessions, costing approximately £35,000 of consultant time. If only two stations were used, this would result in a saving of at least £10,000 (29%) (Table 4).

© 2012 RCPE

TABLE 1 Descriptive statistics

	Station 1		Station 2		Station 3	
	Structured interview	**Suitability**	**Case-based discussion**	**Comm-unication**	**Ethical scenario**	**Profession-alism**
Number of candidates	183	183	183	183	183	183
Minimum score	2.0	1.0	1.0	1.0	1.0	2.0
25% percentile	3.0	3.0	3.0	3.0	3.0	3.0
Median score	4.0	4.0	4.0	4.0	4.0	4.0
75% percentile	5.0	5.0	5.0	5.0	4.0	5.0
Maxiumum score	5.0	5.0	5.0	5.0	5.0	5.0
Mean score	**3.9**	**3.9**	**3.6**	**3.9**	**3.7**	**3.9**
Standard deviation	0.92	0.96	1.1	0.92	1.0	1.0
Standard error	0.068	0.071	0.079	0.068	00.075	0.075
Lower 95% confidence interval (CI) of mean	3.7	3.8	3.5	3.8	3.5	3.7
Upper 95% confidence interval (CI) of mean	4.0	4.1	3.8	4.1	3.8	4.0

TABLE 2 Correlation of stations with each other

	Station 1		Station 2		Station 3	
Parameter	**Structured interview**	**Suitability**	**Case-based discussion**	**Comm-unication**	**Ethical scenario**	**Profession-alism**
n	183	183	183	183	183	183
Spearman (r) correlation	0.765	0.830	0.857	0.759	0.792	0.823
p value	p<0.0001	p<0.0001	p<0.0001	p<0.0001	p<0.0001	p<0.0001

TABLE 3 Correlation between stations

	Station 1		Station 2		Station 3	
Parameter	**Structured interview**	**Suitability**	**Case-based discussion**	**Comm-unication**	**Ethical scenario**	**Profession-alism**
n	183	183	183	183	183	183
Spearman (r) correlation	–0.70	–0.69	–0.69	–0.69	–0.78	–0.74
p value	p<0.0001	p<0.0001	p<0.0001	p<0.0001	p<0.0001	p<0.0001

Assessor survey

The assessor survey received 33 (28%) responses. In the majority of cases, comments were in favour of the new process with many believing that the judgements of the two assessors were often concordant and produced a sensitive and fair system. The majority (87.1%) of assessors felt the purpose of selection was to both assess what the candidate already knows and the candidate's potential to learn. Concerns were expressed regarding the effect of the different amounts of candidates' medical experience on interview performance. Assessors felt the case-based discussion to be the most discriminatory station (Table 5). If only one station were to be used it was felt that it should be the case-based discussion, followed by the structured interview.

Candidate survey

We carried out an analysis of the 137 questionnaires returned from candidates. Candidates perceived all three interview stations to be discriminatory, but thought that the structured interview and clinical stations were more discriminatory than the ethics and professionalism station.

© 2012 RCPE

TABLE 4 Estimation of costs of consultant contribution to selection process

	Three stations	Two stations
Sessions	21	21
Assessors per session	6	4
Assessor sessions	126	84
Extra sessions	21	21
Total	147	105
Consultant pay per year[5]	£125,373	£125,373
Sessions per year	520	520
Consultant cost per session	£241.10	£241.10
Total cost	£35,442	£25,315.50
Number of candidates	183	183
Cost per candidate	£193.67	£138.34

DISCUSSION

All three stations appeared to discriminate for candidate suitability for employment as a CMT trainee and no one station was superior. A reduction in the number of assessment stations from three to any two would not have changed overall candidate ranking i.e. the same doctors would have been offered employment in the same order of preference. Using two stations rather than three would be a more efficient use of resources, potentially saving £10,000 worth of consultants' time. At the same time that data were being collected for this study, pilot work was also undertaken in other deaneries on the impact of using a machine marked test (MMT) rather than a structured application form to shortlist candidates for interview.[10] The study concluded that the MMT was a reliable and efficient method of shortlisting and that an MMT score correlated with subsequent interview performance. An MMT followed by a two-station interview process therefore may prove to be a valid selection process with an efficient use of resources.

The response from assessors and candidates about the 2009 Northern Deanery selection process was very positive. Both groups perceived the assessment stations to be fair and to discriminate for ability, confirming face validity of the selection process. Candidates expressed a preference for the structured interview or case-based discussion, while assessors felt that the case-based discussion was most appropriate for assessing candidate ability. It is of interest that candidates and assessors both perceived the ethical scenario as the least useful station in terms of discrimination, while the data suggests that the candidate score for this station most closely correlated with overall ranking. The reason for this discrepancy is unclear. However, acknowledging the preferences of those taking part in the selection process is important in terms of encouraging the ongoing candidate applications for CMT and the continued involvement of interested clinicians in the selection process.

TABLE 5 Assessor preference if only one station was to be used

	2009
	n=40
Structured interview	7 (23.3%)
Case-based discussion	18 (60%)
Ethics and communication	5 (16.7%)

This is a limited study conducted in a single Deanery, in a single year and can be considered at best a pilot study. Data from the 2007 and 2008 Northern Deanery selection process were reviewed, however, due to the differences in the selection process, the data were not comparable. A further limitation is the disappointing response rate to the assessor survey (28%). Conclusions drawn on the basis of this sample may not accurately reflect the views of assessors as a whole. The survey was conducted after the interviews were completed – a higher response rate would have been obtained if the questionnaires were completed on the selection days but this was not logistically possible. Now that CMT selection is a national process it would be useful to repeat this work in a range of assessment centres to confirm the findings before initiating any changes to the recruitment process.

Perhaps the greatest limitation of this local evaluation is the lack of long-term data available to assess the external predictive validity of the current selection process. At this time it is not known whether those selected for CMT in this way go on to realise their full potential during their training time or to perform well on the job. It is unclear for example whether a single simulated ethics and professionalism station that requires a candidate to break news empathetically means that the trainees will behave this way on a daily basis. Patients themselves might highlight different desirable candidate criteria than those we have chosen for the CMT selection process. Evaluation of the predictive validity would require collection of follow-up data on candidate performance as CMT trainees and subsequent correlation of this with the original selection data. The inherent difficulties of attempting to assess 'on-the-job' performance goes some way to explain the paucity of postgraduate validation studies, in comparison to the more extensive research correlating selection criteria for medical school with subsequent performance as an undergraduate student.[11]

© 2012 RCPE

CONCLUSION

Data from this study indicates that the process of CMT selection in the Northern Deanery could be made more efficient by reducing the number of stations from three to two, without changing overall candidate rankings. If the assumption is that two separate scores from two separate stations will all correlate with each other and the final score/ranking, then it is difficult to justify the use of additional NHS resource to extend the selection process by what appears to be an unnecessary station. This was ultimately a selection process for candidate potential as a CMT trainee and a consultant physician, rather than a clinical assessment or examination. Although this evaluation cannot claim to confirm predictive validity nor construct validity for academic attainment, we have presented evidence suggesting the selection process has concurrent and face validity. Our results suggest that implementing similar selection models nationally for CMT and other specialities could improve the efficiency of the process, although ongoing data collection is needed to determine whether current models of postgraduate selection correlate with future performance. With the evolution of national selection schemes we have an opportunity to look critically at how we select the specialists of the future and to develop an evidence base for our methods, ideally with data from different deaneries over many years, bearing in mind the economic pressures on the NHS and this potential cost saving.

REFERENCES

1 Brown CA, Lilford RJ. Selecting medical students. *BMJ* 2008; 336:786. http://dx.doi.org/10.1136/bmj.39517.679977.80

2 Donaldson L. *Unfinished business. Proposals for the reform of the Senior House Officer grade. A paper for consultation* [Internet]. London: NHS; 2002 [cited 2012 April 18]. Available from: http://217.154.121.42/doh/Docs/Unfinished-Business.pdf

3 Tooke J. *Aspiring to excellence. Findings and final recommendations of the Independent Inquiry into Modernising Medical Careers.* London: MMC Inquiry; 2008.

4 Revill J. Inquiry raps doctor training reform. *The Observer*. 2007 Oct 7.

5 Joint Royal Colleges of Physicians Training Board. *CT1 recruitment* [Internet]. London: Royal College of Physicians; 2012 [cited 2011 May]. Available from: http://www.cmtrecruitment.org.uk

6 Patterson F, Ferguson E. Selection for medical education and training. In: Swanwick T, editor. *Understanding medical education. Evidence, theory and practice.* Edinburgh: Wiley-Blackwell; 2010. p. 352–65. http://dx.doi.org/10.1002/9781444320282.ch24

7 Mitchison H. Assessment centres for core medical training: how do assessors feel this compares with the traditional interview? *Clin Med* 2009; 9:147–50.

8 Berkin L, Burr B. Recruiting physicians to core medical training: how should it be done? *Clin Med* 2009; 9: 410–11.

9 Survey Monkey [Internet]. [cited 2011 Oct]. Available from: http://try.surveymonkey.com

10 Patterson F, Carr V, Zibarras L et al. New machine-marked tests for selection into core medical training: evidence from two validation studies. *Clin Med* 2009; 9: 417–20.

11 Ferguson E, James D, Madeley L. Factors associated with success in medical school: systematic review of the literature. *BMJ* 2002; 324: 952–7. http://dx.doi.org/10.1136/bmj.324.7343.952

© 2012 RCPE