

Content validity of a clinical problem solving test for use in recruitment to the acute specialties

Gemma Crossingham, Thomas Gale, Martin Roberts, Alison Carr, Jeremy Langton and Ian Anderson

ABSTRACT – Clinical problem solving tests (CPSTs) have been shown to be reliable and valid for recruitment to general practice (GP) training programmes. This article presents the results from a Department of Health-funded pilot into the use of a CPST designed for recruitment to the acute specialties (AS). The pilot paper consisted of 99 items from the validated GP question bank and 40 new items aimed specifically at topics of relevance to AS training. The CPST successfully differentiated between applicants. The overall test and the GP section showed high internal reliability, whereas the AS pilot section performed less well. A detailed item analysis revealed that the AS pilot items were, on average, more difficult and of poorer quality than the GP items. Important issues that need to be addressed in the early development phase of a test used for high stakes selection to specialty training programmes are discussed.

KEY WORDS: acute specialties, clinical problem solving test, content validity, machine marked test, postgraduate, specialty selection

The 2008 Tooke Report *Aspiring to excellence*¹ made the following comment in regard to the shortlisting process for recruitment to specialty training posts:

There would be considerable attractions in having a scheme which was both more accurate and less labour intensive. Successful models for shortlisting include the UK GP [general practice] selection system and the US system, both of which are based on scores in applied knowledge tests. These have advantages in being able to identify weak applicants...Such an approach deserves further evaluation.

Subsequent research has shown the machine marked tests (MMTs) used in GP selection are as valid and reliable as the current standard shortlisting method of scoring application form questions.² In light of this, this article presents the preliminary results from an ongoing Department of Health-funded pilot

study of the development of a clinical problem solving test (CPST) tailored to the acute medical specialties, looking specifically at its reliability, content validity and face validity.

The initial pilot test

The CPST paper consisted of 139 items mapped onto the foundation programme curriculum: 99 taken from the GP bank of validated items and 40 newly written items aimed specifically at topics of relevance to acute specialty (AS) training. The tried and tested GP items provided a useful benchmark against which to assess the performance of the newly developed AS items. The new items were written by a group of subject matter experts, including consultants in acute medicine, anaesthesia, emergency medicine and intensive care medicine.

Applicants attending selection centres in 2008 at the South West Peninsula Deanery for CT1 training posts in anaesthesia, core medical training and the acute care common stem (ACCS) were invited to sit the first pilot CPST. Participation (or not) did not influence selection outcome in any way and participants consented to the linking of test results with other personal data. The paper was administered on the day of interview and an overall response rate of 74% (125 of 169 applicants) was achieved.

Reliability of the test

The distribution of test scores was approximately normal in each section (GP and AS), indicating an absence of ceiling or floor effects and showing that the test has the potential to differentiate between candidates. Both the overall test and the GP item section showed very high internal reliability (Cronbach's $\alpha=0.90$ and 0.92 respectively). The AS pilot section performed less well ($\alpha=0.43$), though this is partly attributable to the smaller number of items used. When corrected for test length (Spearman–Brown), the expected reliability of a 99-item test of equivalent AS items would be 0.65 , still below the 0.9 threshold recommended for high stakes assessment.³

A detailed item analysis revealed that the AS pilot items were on the whole more difficult (mean item facility = 0.50 v 0.61) and of poorer quality (mean item partial = 0.09 v 0.29) than the GP items. This is not an entirely unexpected finding; the GP items had been refined over many years of development whereas the AS specific items were all being administered for the first time having been written on a tight timescale before the 2008 recruitment round.

Gemma Crossingham,¹ ST5 in anaesthesia; Thomas Gale,¹ consultant anaesthetist; Martin Roberts,² research fellow; Alison Carr,¹ consultant anaesthetist; Jeremy Langton,¹ consultant anaesthetist; Ian Anderson,¹ consultant anaesthetist

¹Plymouth Hospitals NHS Trust, Plymouth

²Institute of Clinical Education, Peninsula College of Medicine and Dentistry, Plymouth

Table 1. Applicant ratings of selection centre stations and machine marked tests (MMTs) for ‘fairness’, ‘opportunity to demonstrate ability’ and ‘relevance’.

| Applicant feedback ratings (n=43) | Fairness | | Opportunity | | Relevance | |
|--|----------|------|-------------|------|-----------|------|
| | Mean | SD | Mean | SD | Mean | SD |
| Structured interview | 4.21 | 0.78 | 4.07 | 0.67 | 4.09 | 0.75 |
| Portfolio station | 4.29 | 0.72 | 4.15 | 0.79 | 4.17 | 0.70 |
| Presentation station | 4.01 | 0.75 | 4.02 | 0.80 | 3.93 | 0.77 |
| Simulation station | 4.05 | 0.79 | 4.14 | 0.80 | 4.21 | 0.91 |
| Role-play station | 4.46 | 0.78 | 4.38 | 0.71 | 4.21 | 0.78 |
| Telephone station | 4.25 | 0.68 | 4.08 | 0.78 | 4.17 | 0.82 |
| MMT | 3.90 | 0.89 | 3.90 | 0.94 | 3.33 | 0.91 |
| P-value for Kruskal-Wallis test of differences in ratings | 0.154 | | 0.223 | | <0.001 | |

Applicant feedback

In order to assess the face validity of the MMT, applicants completed a feedback questionnaire to evaluate the MMT alongside six other selection centre stations used on the day of interview. Applicants were asked to rate the selection tools for relevance to selection, fairness and opportunity to demonstrate ability using a five-point rating scale (1 = poor; 2 = borderline; 3 = satisfactory; 4 = good; 5 = excellent). Although the MMT was less well regarded than the other selection stations in all three aspects, only the ratings for relevance were significantly lower than the other selection stations (Table 1).

Relevance of the items

It was postulated that the significantly lower score for ‘relevance to selection’ might be due to the use of primary care-related questions in an AS population. To investigate this further, 59 trainees and consultants in anaesthesia, core medical training and AS were asked to rate a random sample of items taken from the pilot CPST for ‘appropriateness’ to selection for specialty training on a scale of 1 (totally inappropriate) to 4 (entirely

appropriate). The random sample was stratified by question source to contain 15 items from the GP section of the test and 15 from the AS section. Respondents were blinded to the source of the items, which were presented in random order. As shown in Fig 1 the AS items were largely regarded as appropriate, whereas the GP items were mainly felt to be inappropriate for selection to specialty training. Average ratings for the AS items (mean=3.43, SD=0.25) were higher than those for the GP items (mean=2.09, SD=0.41) and the difference was statistically significant (t-test=10.68, df=23, p<0.001).

Discussion

The use of an MMT potentially enables standardisation of short-listing processes, is time-effective, cost-effective and can predict subsequent success at postgraduate examinations.^{4,5} A nationally administered MMT is established for recruitment to UK training posts for general practice and has been used for core medical training with evidence of good reliability and validity when compared to standard shortlisting and interview scores.^{2,6}

Similarly, validated CPST questions from the GP question bank perform well in terms of reliability and item quality in an AS population. However, the content validity of GP items was significantly lower than items written specifically for the acute specialties within this population. Both the GP question bank and the AS items used in this pilot are mapped onto the foundation programme curriculum⁷ and National Person Specifications for specialty training posts.⁸ However, it is evident that different specialty recruitment processes may focus on various aspects of foundation experience. Content validity is an important consideration when introducing a new method of recruitment and justifies robust test specification on nationally agreed blueprints.

Face validity refers, not to what the test actually measures but, to what it appears to measure and pertains to whether the test ‘looks valid’ to the applicants who take it. Applicants rated the MMT lower than the other selection centre stations for relevance at the point of interview. This may be due to the fact that applicants do not like sitting a paper-based test at the point of

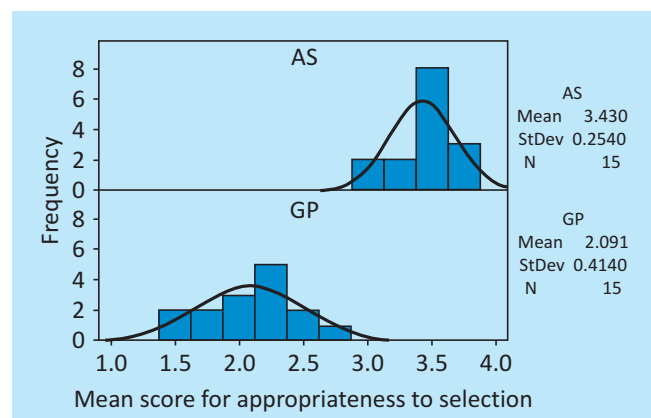


Fig 1. Histograms of mean scores for appropriateness to selection awarded to acute specialty (AS) and general practice (GP) questions split by question source (n=59×30=1,770).

interview but may be something that can be improved with further refinement of test specification and items used. The study aims to evaluate the MMT as it would be used in practice: at the point of application rather than at a future interview.

Recent evidence suggests that the combination of a situational judgement test (SJT) with a CPST can maximise validity and efficiency.² SJTs target non-clinical domains and assess professional behaviour and attitudes that have been found to correlate with subsequent workplace performance.⁹ An SJT based on complex job analysis, tailored to the assessment of professionalism in an acute specialty population is currently under development.

Important issues that need to be addressed in the early development phase of a test used for high stakes assessment for selection to specialty training programmes have been described. Having tested the reliability of a generic CPST in the AS population in this pilot, the next phase will involve test refinement for content and item quality with expansion of the specialty specific item pool. Subsequent studies will be extended to include units of application throughout the UK to further test the refined items on a wider demographic sample and investigate the construct validity of the MMT in relation to current shortlisting and interview processes.

Acknowledgements

We would like to thank Dr Simon Plint and the GP National Recruitment Office for contributing items for the test. Professor Fiona Patterson and Dr Victoria Carr (Work Psychology Group) for their assistance with data analysis. Dr Peter Riou (emergency medicine consultant) and Dr Nick Withers (consultant physician) made significant input to the acute specialties question bank. Mr Neil Squires (specialty recruitment manager), Dr Peter Davies (training programme director, anaesthesia) and members of the South West Peninsula Deanery team were invaluable in implementing the test. This project was run by the

Anaesthesia Recruitment Validation Group from the South West Peninsula Deanery with significant support from the Directorate of Anaesthesia, Derriford Hospital. Study supported by Department of Health funding.

References

- 1 Tooke J. *Aspiring to excellence: findings and recommendations of the Independent Inquiry into Modernising Medical Careers*. London: MMC Inquiry, 2008.
- 2 Patterson F, Baron H, Carr V, Plint S, Lane P. Evaluation of three short-listing methodologies for selection into postgraduate training in general practice. *Med Educ* 2009;43:50–7.
- 3 Nunnally JC, Bernstein IH. *Psychometric theory*, 3rd edn. New York: McGraw-Hill, 1994.
- 4 Carmichael KDMD, Westmoreland JBMD, Thomas JAMD, Patterson RMP. Relation of residency selection factors to subsequent orthopaedic in-training examination performance. *Southern Med J* 2005;98:528–32.
- 5 Matveevskii AS, Gravenstein N. Role of simulators, educational programs, and nontechnical skills in anesthesia resident selection, education, and competency assessment. *J Crit Care* 2008;23:167–72.
- 6 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.
- 7 Department of Health. *The Foundation Programme Curriculum*. London: Department of Health, 2007.
- 8 Medical Specialty Training (England) Person Specifications. www.mmc.nhs.uk/specialty_training_2010/specialty_training_2009/person_specifications.aspx#themed
- 9 McDaniel MA, Morgeson FP, Finnegan EB, Campion MA, Braverman EP. Use of situational judgement tests to predict job performance: a clarification of the literature. *J Appl Psychol* 2001;86:730–40.

**Address for correspondence: Dr G Crossingham, Department of Anaesthesia, Plymouth Hospitals NHS Trust, Derriford Road, Plymouth, Devon PL6 8DH.
Email: gemma.crossingham@phnt.swest.nhs.uk**