

New machine-marked tests for selection into core medical training: evidence from two validation studies

Fiona Patterson, Victoria Carr, Lara Zibarras, Bill Burr, Liz Berkin, Simon Plint, Bill Irish and Simon Gregory

ABSTRACT – This study examined whether two machine-marked tests (MMTs; a clinical problem-solving test and situational judgement test), previously validated for selection into UK general practice (GP) training, could provide a valid methodology for shortlisting into core medical training (CMT). A longitudinal design was used to examine the MMTs' psychometric properties in CMT samples, and correlations between MMT scores and CMT interview outcomes. Independent samples from two years were used: in 2008, a retrospective analysis was conducted ($n=1,711$), while in 2009, CMT applicants completed the MMTs for evaluation purposes ($n=2,265$). Both MMTs showed good reliability in CMT samples, similar to GP samples. Both MMTs were good predictors of CMT interview performance ($r=0.56$, $p<0.001$ in 2008; $r=0.61$, $p<0.001$ in 2009) and offered incremental validity over the current shortlisting process. The GP MMTs offer an appropriate measurement methodology for selection into CMT, representing a significant innovation for selection methodology.

KEY WORDS: interviews, machine-marked tests, selection, shortlisting, validity

Introduction

Selection into postgraduate training has been a widely debated topic, especially in the UK. Recently, the selection methodology used for entry into UK general practice (GP) has demonstrated good reliability and validity evidence.¹ The methodology includes two invigilated, machine-marked tests (MMTs):

- a clinical problem solving test (CPS) comprising questions that require applicants to apply clinical knowledge to solve problems reflecting diagnostic processes or develop management strategies for patients
- a *situational judgement test* (SJT) where applicants are presented with written depictions of professional dilemmas

^{1,2}Fiona Patterson, Professor of Organisational Psychology; ²Victoria Carr, Consultant Psychometrician; ^{1,2}Lara Zibarras, Research Fellow; ³Bill Burr, Medical Director; ³Liz Berkin, Associate Director; **Simon Plint**, former Chair, GP National Recruitment Office; GP Dean NHS Education South Central; **Bill Irish**, Head, School of Primary Care, Severn Deanery; **Simon Gregory**, Postgraduate Dean, East of England Multi-professional Deanery, UK; former Chair, GP National Recruitment Office

¹City University, London; ²Work Psychology Group; ³Joint Royal Colleges of Physicians Training Board, Royal College of Physicians, London

they may encounter at work and are asked to identify an appropriate response from a list of alternatives.

This paper describes an evaluation study to examine whether these MMTs can provide a valid selection methodology for shortlisting into core medical training (CMT).

Any new selection methodology must satisfy exacting psychometric criteria such as reliability, validity and fairness.^{2–4} This paper therefore reports the psychometric properties of the CPS and SJT for independent samples of CMT applicants in two consecutive years (2008 and 2009) and examines whether scores in these MMTs (time 1) predict performance in the CMT selection interview (time 2) approximately one month later. For the 2009 sample, MMT marks were also compared with the current shortlisting process, based on scores awarded to sections of a structured application form. The objectives of this paper were to explore the reliability, and predictive validity of the CPS and SJT for applicants to CMT. In particular, the aim was to explore any potential gains in effectiveness by using these MMTs compared to current application form scoring procedures. Specifically, we addressed the following research questions:

- 1 Are the psychometric properties of the CPS and SJT robust in the CMT applicant samples?
- 2 Are the CPS and SJT valid methods of selection for CMT applicants? Do they predict subsequent performance in the CMT interview?
- 3 Compared to the current application form-based shortlisting procedure, do the MMTs add incremental validity in predicting outcomes?

Method

Recruitment into general practice and core medical training in the UK

The selection systems for CMT and GP training in the UK are designed to process several thousand applicants nationally in each yearly recruitment round. The GP methodology comprises 3 stages:

- | | |
|---------|---|
| Stage 1 | eligibility checks including demonstration of current evidence of foundation competence |
| Stage 2 | short-listing using the CPS and SJT |
| Stage 3 | a selection centre using work-relevant simulations. |

The CPS paper has 100 items and the SJT 50. The existing CMT selection process comprises:

- | | |
|---------|--------------------|
| Stage 1 | eligibility checks |
|---------|--------------------|

- Stage 2 shortlisting via a structured application form, involving a range of biographical information including achievements, qualifications and a personal statement
- Stage 3 a structured interview (one month after shortlisting), lasting approximately 30 minutes and comprising questions derived from a UK standardised person specification.

Design and sampling

Two validation studies conducted over two consecutive years are reported in this paper, each using a different research design. In 2008, a retrospective evaluation of existing MMT data was conducted. Anonymised data held by the GP National Recruitment Office were used to identify a cohort of applicants who applied to both CMT and GP in the 2008 recruitment round. A total of 8,195 applicants completed the MMTs as part of the selection process for GP training; of which 1,711 had also applied to CMT (this sample represents 49% of total applicants for CMT).

In 2009, a prospective evaluation of the MMTs was conducted alongside live selection. Here, all doctors applying for CMT were asked to complete the MMTs in addition to the live selection process. Applicants were made aware that this was a pilot, and that the MMT marks would not be known by recruiters when the selection decisions were made. Overall, a total of 2,265 CMT applicants completed the MMTs and data for the full GP sample (n=5,311) provided a comparison sample.

For both 2008 and 2009, performance in the CMT interview process provided an outcome measure. In 2009, performance in the CMT shortlisting process was also explored. The reliability of each MMT was calculated using Cronbach's α coefficient. Table 1 details the demographic details and sizes of the two independent CMT samples for 2008 and 2009, and the associated GP comparison samples. Generally, the CMT samples are comparable to their associated comparison group in terms of demographics.

Results

Are the psychometric properties of the CPS and SJT robust in the CMT applicant samples?

Table 2 gives descriptive statistics for the CPS and SJT for the 2008 and 2009 CMT applicant samples, with results contrasted against the relevant comparison sample. Results indicate that both MMTs had good reliability for the CMT sample in both 2008 (CPS $\alpha=0.89$, SJT $\alpha=0.80$) and 2009 (CPS $\alpha=0.85$, SJT $\alpha=0.85$). Score distributions were close to normal showing that the tests were capable of differentiating between applicants. MMT marks for each CMT sample (2008 and 2009) were similar to the relevant GP comparison sample.

The correlation between CPS and SJT scores within the CMT sample in 2008 (n=1,710) was $r=0.45$; in 2009 (n=2,264) the correlation between the two test scores was $r=0.54$. These results are similar to the correlations between the CPS and SJT for GP comparison samples in both 2008 ($r=0.51$) and 2009 ($r=0.53$). The size of correlation between the two MMTs (CPS and SJT) for all samples presented in the study indicates that the tests have both common and independent variance.

Do the CPS and SJT predict subsequent performance in the CMT interview?

Correlations between MMT marks and CMT interview scores were calculated to examine whether the tests predict performance in the CMT selection process. Both the 2008 and 2009 samples are based on applications rather than individual applicants. In 2008, data were available for 837 applications. In 2008, deaneries used different CMT interview processes, therefore data from each deanery were analysed separately. Deaneries with less than 30 cases were not included in correlations due to small sample size. The average uncorrected correlation with CMT interview scores was $r=0.43$ for the CPS; $r=0.52$ for the SJT, and $r=0.56$ for MMTs combined. In 2009, all deaneries used the same CMT interview process, allowing overall correlations to be calculated. In the 2009 sample (n=3,231), the

Table 1. Sample details and breakdown.

	n	Mean age	Women (%)	Men (%)	White (%)	Asian (%)	Black (%)	Mixed (%)	Other (%)
<i>2008 samples</i>									
CMT sample that completed MMTs	1,711	30.8	46	53	23	62	6	3	6
GP comparison sample	8,195	31.6	48	50	33	51	7	3	5
<i>2009 samples</i>									
CMT sample that completed MMTs	2,265	28.4	60	40	45	36	6	3	10
GP comparison sample	5,311	30.2	56	44	44	42	7	1	6

Note: % may not total 100 due to missing data.

CMT = core medical training; GP = general practice; MMT = machine-marked tests.

overall uncorrected correlation with CMT interview scores was $r=0.54$ for the CPS; $r=0.53$ for the SJT, and $r=0.61$ for MMTs combined (all $p<0.001$). This represents a strong level of prediction for any selection methodology,⁵ similar to that of the comparison samples as a whole.¹ This is an important finding, indicating that the MMTs are robust instruments for shortlisting purposes.

Compared to the current application form-based 'traditional' shortlisting procedure, do the MMTs add incremental validity in predicting outcomes?

For the 2008 sample, hierarchical regression analyses showed that in almost all cases, the SJT offered significant incremental validity over the CPS in predicting interview scores, on average accounting for an additional 15% of the variance. In around half of cases, the CPS also offered incremental validity over the SJT, on average accounting for an additional 6% of the variance. This indicates that the SJT is the best single predictor of CMT interview scores but that the CPS and SJT in combination provide the strongest evidence of predictive validity.

For the 2009 sample, the overall uncorrected correlation with traditional CMT shortlisting scores was $r=0.40$ for the CPS, $r=0.34$ for the SJT and $r=0.42$ for both tests combined (all $p<0.001$). The correlation between CMT shortlisting scores and CMT interview scores was $r=0.46$ ($p<0.001$), indicating that the traditional shortlisting methodology is valid and provides good prediction of outcomes at interview. Both the CPS and SJT offered significant incremental validity over each other in predicting CMT interview scores (accounting for an additional 9% and 8% of the variance respectively).

When comparing the MMTs with the traditional CMT shortlisting methodology, a multiple stepwise regression confirmed that the MMTs contributed most to the prediction of CMT interview scores, over the traditional CMT shortlisting process. The MMTs offered significant incremental validity over the traditional shortlisting process: the CPS accounted for an additional 16%; the SJT an additional 17%; and both tests

combined, an additional 22% of the variance in interview scores. This is a high level of incremental validity indicating that the MMTs add significantly to the prediction of interview outcomes over the traditional application form shortlisting process alone. From a cost-efficiency perspective, the MMTs are likely to offer significant advantages compared to traditional shortlisting processes.

Discussion

Overall, in this study we set out to determine:

- whether the psychometric properties of the CPS and SJT were robust in independent CMT applicant samples over two consecutive years
- whether the CPS and SJT offer valid methods of selection for CMT applicants by predicting subsequent performance in the CMT interview
- whether the CPS and SJT offer incremental validity over existing shortlisting measures in predicting interview outcomes.

The findings provide good evidence to suggest that tests such as those outlined in this paper show promise as a selection methodology for CMT in the UK. Both MMTs showed sufficient reliability within the CMT sample (all Cronbach's $\alpha \geq 0.80$) and score distributions similar to the GP comparison sample. Further, both the CPS and SJT are valid tests because in each independent sample over two consecutive years (2008 and 2009) they are strong predictors of subsequent performance in CMT interviews. In addition, findings from 2009 clearly indicate that both MMTs have substantial incremental validity over the current CMT shortlisting process. In general, the results demonstrate that the MMTs add significant value in predicting interview outcomes.^{4,6,7}

The 2008 CMT sample comprised only a subset of applicants – those who applied to both CMT and GP. It could be argued that this sample may not fully represent the CMT applicant population as a whole, so that results may not readily generalise to the applicants not included in this sample. However, the 2009 sample involved the whole CMT applicant population and

Table 2. Clinical problem solving (CPS) and situational judgement test (SJT) results.

	n	CPS			SJT		
		Mean score (standard)*	Standard deviation	Reliability (α)	Mean score (standard)*	Standard deviation	Reliability (α)
<i>2008 samples</i>							
CMT	1,711	255	36	0.89	249	37	0.80
GP comparison	8,195	250	40	0.91	253	40	0.83
<i>2009 samples</i>							
CMT	2,265	255	38	0.85	250	40	0.85
GP comparison	5,311	250	40	0.86	250	40	0.85

*Scores on each test are standardised to a scale with a mean of 250 and a standard deviation of 40 (based on the distribution of scores for the general practice (GP) sample). CMT = core medical training.

F Patterson, V Carr, L Zibarras, B Burr, L Berkin, S Plint, B Irish and S Gregory

results obtained were similar over the two consecutive years. Replicating the results in this way gives further confidence that the MMTs are not only psychometrically robust, but are also able to predict performance in the CMT interviews.

Based on the available evidence therefore, results indicate that the MMTs reported in this paper are an appropriate measurement methodology for shortlisting in CMT. These MMTs can provide a standardised process that could substantially increase the cost-effectiveness and utility of the selection procedure once an initial development phase has been completed.⁴ Since the tests are machine marked, the methodology offers significant advantages over traditional shortlisting approaches in providing increased efficiencies and resource savings, which would benefit the CMT selection procedure. For the future, these MMTs could be evaluated for use as a single shortlisting methodology for several medical specialties, thereby offering multiplicative efficiency savings.

To explore the use of the MMTs in other specialties, job analysis studies are required to ensure the selection methods used are targeting appropriate criteria for entry into specialty training. Job analysis studies address issues such as content validity, and outputs will ensure that the selection methodology is geared towards the prediction of longer-term outcomes (for example training progression).^{8,9} Job analysis studies can inform development of a test specification, which can be modelled with appropriate subject matter experts to define test content and to review standard setting procedures for the tests. In accordance with best practice, a long-term validation study will ensure that the tests predict in-training performance. Further research should explore applicant reactions to the selection process so that perceptions of fairness and justice can be monitored over time.¹⁰

Competing interests

FP, VC and LZ conduct work for the Work Psychology Group who advises the Department of Health (DH) in the UK on selection and recruitment issues. SP is seconded to the DH's Modernising Medical Careers team, but the views expressed are personal.

Authorship and contributorship

FP and VC conceived of the methodology and measures in the study. FP, VC and LZ analysed and interpreted the data, and wrote the paper.

BB, LB, SP, SG and BI conceived of the original study, contributed to the overall study design, organised data collection and interpreted the data. All authors commented on the final version of the paper.

Acknowledgements

Mrs Gai Evans is acknowledged for her significant contribution to data collection via the GP National Recruitment Office, and Ms Helen Kelly for coordinating the data collection from core medical training. Ms Helen Baron and all the item writers are acknowledged for their contribution to the development of the tests. The various contributors from all locations helping in data collection are acknowledged.

References

- 1 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.
- 2 Patterson F, Ferguson E, Norfolk T, Lane P. A new selection system to recruit GP registrars: Preliminary findings from a validation study. *BMJ* 2005;330:711–4.
- 3 Robertson IT, Smith M. Personnel selection. *J Occup Organ Psych* 2001;74:441–72.
- 4 Schmidt FL, Hunter JE. The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychol Bull* 1998;124:262–74.
- 5 Patterson F, Ferguson E. Selection into medical education and training. *Understanding Medical Education; ASME Publications* 2007.
- 6 Hunsley J, Meyer, GJ. The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical Issues. *Psychol Assessment* 2003;15:446–55.
- 7 McDaniel MA, Morgeson FP, Finnegan EB, Campion MA, Braverman EP. Use of Situational Judgment Tests to predict job performance: A clarification of the literature. *J App Psychol* 2001;86:730–40.
- 8 Arnold J, Silvester J, Patterson F *et al.* *Work psychology: understanding human behaviour in the workplace*. London: FT Prentice Hall, 2004.
- 9 Borman WC, Hanson MA, Hedge JW. Personnel selection. *Annu Rev Psychol* 1997;48:299–337.
- 10 Patterson F, Zibarras L, Carr V, Irish B, Gregory, S. Evaluating candidate reactions to selection practices. *Med Ed* (submitted).

**Address for correspondence: Professor F Patterson,
City University, Northampton Square, London EC1V 0HB.
Email: f.patterson@city.ac.uk**