
This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: http://openaccess.city.ac.uk/12507/

Link to published version:

Copyright and reuse: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.
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 – The objective of this study was to examine whether two machine-marked tests (MMTs), a clinical problem-solving test (CPS) and a situational judgement test (SJT; focusing on professional dilemmas), previously validated for selection into general practice (GP) training in the UK, could provide a valid selection methodology for shortlisting into core medical training (CMT). An exploratory longitudinal design was used to examine the psychometric properties of the MMTs for CMT applicants, and the correlation between MMT scores (time 1) and subsequent CMT interview outcomes (time 2). Independent samples from two consecutive years were used: in 2008 a retrospective analysis of data was used, while in 2009, doctors applying for CMT were asked to complete the MMTs for evaluation purposes. In 2008, a total of 1,711 doctors applied to both CMT and GP training and completed the MMTs. In 2009, a total of 2,265 doctors who applied for CMT completed the MMTs for evaluation purposes. The main outcome measure was the CMT applicants’ interview score. Both the CPS and SJT had good reliability and score distributions for 2008 and 2009 CMT samples, similar to the GP comparison samples. The MMTs were good predictors of performance in the CMT interviews ($r=0.56$, $p<0.001$ in 2008, and $r=0.61$, $p<0.001$ in 2009 for both MMTs combined) 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 developing selection methodology for CMT. Longer-term studies should be undertaken to assess the validity of all selection techniques used, in terms of CMT training outcome.

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 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, item quality 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:
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 tests 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. In 2008, data were available for 837 applicants. It should be noted that in 2008,
new machine-marked tests for selection into core medical training

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

<table>
<thead>
<tr>
<th></th>
<th>CPS</th>
<th>SJT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean score (standard)*</td>
</tr>
<tr>
<td>2008 samples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMT</td>
<td>1,711</td>
<td>255</td>
</tr>
<tr>
<td>GP comparison</td>
<td>8,195</td>
<td>250</td>
</tr>
<tr>
<td>2009 samples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMT</td>
<td>2,265</td>
<td>255</td>
</tr>
<tr>
<td>GP comparison</td>
<td>5,311</td>
<td>250</td>
</tr>
</tbody>
</table>

*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.

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 α ≥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

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. In 2009, all deaneries used the same CMT interview process, allowing overall correlations to be calculated. Both the 2008 and 2009 samples are based on applications rather than individual applicants. 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 the 2009 sample (n=3,231), the 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, particularly 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).
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


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

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.