The predictive validity of a situational judgement test, a clinical problem solving test and the core medical training selection methods for performance in specialty training

Authors: Fiona Patterson,[A] Safiatu Lopes,[B] Stephen Harding,[C] Emma Vaux,[D] Liz Berkin[E] and David Black[F]

ABSTRACT

The aim of this study was to follow up a sample of physicians who began core medical training (CMT) in 2009. This paper examines the long-term validity of CMT and GP selection methods in predicting performance in the Membership of Royal College of Physicians (MRCP(UK)) examinations. We performed a longitudinal study, examining the extent to which the GP and CMT selection methods (T1) predict performance in the MRCP(UK) examinations (T2). A total of 2,569 applicants from 2008–09 who completed CMT and GP selection methods were included in the study. Looking at MRCP(UK) part 1, part 2 written and PACES scores, both CMT and GP selection methods show evidence of predictive validity for the outcome variables, and hierarchical regressions show the GP methods add significant value to the CMT selection process. CMT selection methods predict performance in important outcomes and have good evidence of validity; the GP methods may have an additional role alongside the CMT selection methods.

KEYWORDS: Assessment, CMT, core medical training, CT1, interviews, machine-marked tests, selection, validity

Introduction

In the UK, up to 3,000 junior doctors apply annually to core medical training (CMT), through a nationally coordinated process, to train as physicians.[1] To ensure that individuals selected for training will become competent in practice, selection assessments must be valid, fair and legally defensible.[2] In particular, establishing the predictive validity of a selection method is central to understanding the extent to which a method can predict applicants' future performance.[3] This

Authors: [A]director, Work Psychology Group, Derby, UK; [B]consultant psychologist, Work Psychology Group, Derby, UK; [C]specialty recruitment office team manager, Joint Royal Colleges of Physicians Training Board, London, UK; [D]consultant nephrologist Royal Berkshire NHS Foundation Trust, Reading, UK; [E]consultant cardiologist, Leeds Teaching Hospitals Trust, Leeds, UK; [F]medical director, Joint Royal Colleges of Physicians Training Board, London, UK

ensures that selection methods identify the best person for the role.

To date, there has been no published evidence of the longer-term, predictive validity of the CMT selection process. However, a 2009 study explored the GP machine-marked tests (MMTs) for shortlisting into CMT.[4] Patterson *et al* included two invigilated MMTs:

1 a clinical problem solving test (CPS), designed to measure applicants' ability to apply clinical knowledge in a relevant context and make clinical decisions in practice

2 a situational judgement test (SJT), where applicants were presented with text-based scenarios of professional dilemmas they may encounter at work and asked to identify an appropriate response from a list of alternatives.

This 2009 study demonstrated that the GP MMTs were reliable and predictive of subsequent performance in CMT selection interviews, suggesting that the MMTs may be a useful selection methodology for CMT in the UK. However, to date no validation work has been conducted to further substantiate these findings, even though the 2009 paper suggested that future research studies should explore the prediction of longer-term outcomes, including progression during training.

This study expands on Patterson and colleagues' 2009 research by following up the same sample of physicians who began training in 2009 to examine the validity of the selection methods in predicting performance in the Membership of Royal College of Physicians (MRCP(UK)) diploma examinations. It is important to explore this longitudinal predictive validation data to assess the reliability and validity of the CMT selection process for selecting trainee doctors who are likely to be successful during CMT.[5] In this paper, we present research that explores the extent to which the various selection methods predict important longer-term outcomes. The selection methods explored in this study included the MMTs along with the CMT selection methods.

After successful completion of the UK Foundation Programme (or comparable training experience), doctors are selected into CMT through a standardised and nationally coordinated process, including shortlisting (on the basis of achievements and qualifications) and an interview process. The interview includes three stations: Station 1 measures suitability and commitment to the specialty (and is also a review of a

candidate's portfolio); Station 2 measures communication skills during handling of a clinical scenario; and Station 3 measures professionalism and governance and also includes an exercise for reviewing responses to an ethical scenario. Once appointed, during the course of CMT, trainees undertake the MRCP(UK) examination and must pass all three parts of this exam before they become eligible to enter specialty training at ST3, where they can apply to specialise in one of over 30 different specialties, such as neurology, cardiology and dermatology.

In order to explore the longer-term validity of the MMTs and the CMT selection methods we posed the following research questions:

1. To what extent do the CMT selection methods and the MMTs correlate with subsequent performance in MRCP(UK) examinations?
2. What is the incremental predictive validity of the MMTs over CMT selection methods in predicting performance in MRCP(UK) examinations?

Methods

Design and sampling

The selection data (MMTs and CMT selection methods) for all applicants who applied to both CMT and GP specialty training in 2008/2009 were used, as well as data from CMT applicants who did not apply to GP training, but completed the MMTs as a pilot. In the original study by Patterson *et al*,[4] two validation studies were conducted using different research designs. The 2008 sample was a retrospective evaluation to explore a cohort of applicants who applied to both CMT and GP in the 2008 recruitment round (n=1,711). The 2009 sample was a prospective evaluation of the MMTs conducted alongside live selection; however, applicants knew that this was a pilot and that MMT marks would not influence selection decisions (n=1,265).

Data for all applicants from the 2008/2009 CMT selection process were matched to MMT scores using General Medical Council numbers. There were 2,569 applicants to CMT in 2009, 2,434 of which made a second application to a different training region. In instances where the applicants applied twice, we used their first application scores for the analyses. However, if the applicants accepted an offer resulting from their second application, then these application scores were used instead.

Measures

Predictor variables
MMTs
The CPS and the SJT tests were invigilated tests. The CPS had 87 items and lasted 90 minutes; the SJT had 50 items and lasted 110 minutes. Cronbach's alpha of the CPS paper for the sample used in this study were 0.89 and 0.85 for those who completed the papers in 2008 and 2009, respectively; for the SJT paper this was 0.81 and 0.85 for 2008 and 2009, respectively. These findings are very similar to those found by Patterson *et al*:[4] CPS 0.89, SJT 0.80 in 2008 and CPS 0.85, SJT 0.85 in 2009. These alphas are marginally different because in the present study the sample size is smaller and only includes those whose MMT scores have matched successfully with CMT selection.

CMT selection methods
Automated scores out of 64 marks were produced based on self-assessment of achievements and qualifications for shortlisting candidates. Successful applicants were then invited to a standardised interview day, which included three interview stations measuring aptitude/skill in six areas: suitability and commitment to CT1, achievements to date, clinical skills, communication skills, handling of an ethical scenario, and professionalism and governance. In this study, we used the overall interview score out of 60 marks.

Criterion variables

MRCP(UK) examinations
The MRCP(UK) examination entails three components: parts 1 and 2 are both written exams, the third part is a clinical examination (practical assessment of clinical examination skills – PACES). The MRCP(UK) supplied scores for all MRCP(UK) examinations attempts from 2008 to 2012 (a total of 7,583 attempts). CMT/MMT selection method data and the MRCP(UK) examination data matched successfully for 2,569 applicants.

Final sample

Of the 2,569 applicants, 41.8% were male and 57.6% female (0.7% data missing); 59.1% of applicants trained in UK medical schools, 40.9% in non-UK medical schools. The mean age was 29.7 years (range 24–60 years). 44.5% of applicants described themselves as white (36.1% white British or Irish), 36.3% as Asian, 6.3% as black, 4.6% as Chinese, 3.6% as mixed and 4.6% as other.

Results

Table 1 shows the descriptive statistics for all study variables. All variables were normally distributed, except the PACES score, which showed a slight negative, but non-significant, skew.

Analyses were conducted using SPSS (Version 22.0). Outliers were deleted pairwise within cases prior to analysis in order to maintain sample sizes and not distort the findings. Correlations (Pearson's *r*) between CMT/MMT selection method scores and MRCP(UK) examination scores were calculated to examine the relationship between the selection methods and examination scores (Table 2).[a]

To what extent do the CMT selection methods and the MMTs correlate with subsequent performance on MRCP(UK) examinations?

All predictors were significantly correlated with each other (p<0.001) and were significantly positively correlated with the outcome variables (p<0.001), providing initial evidence of the predictive validity of the selection methods. As might be expected, the strongest association was found between the CPS and both the MRCP part 1 and part 2 (*r*=0.69 and *r*=0.57, respectively; p<0.001), which all assess clinical knowledge. Similarly, for PACES, the strongest association was found with the CMT interview score (*r*=0.45, p<0.001), followed by the SJT (*r*=0.40, p<0.001) as these assessments all focus on non-academic attributes.

Table 1. Descriptive statistics for study 1 variables

Variable	n	Mean	SD	Min	Max
Age (as of 01.01.09), years	2,561	29.74	4.24	24	60
Predictor variables					
CMT shortlisting score	2,569	29.46	11.98	0.00	75.00
CMT total interview score	1,974	45.65	8.00	18.00	60.00
CPS score	2,275	255.44	37.55	123.00	334.00
SJT score	2,263	251.61	37.41	113.00	329.00
Criterion variables					
MRCP(UK) part 1 score[*]	1,693	−4.36	12.31	−40.80	27.53
MRCP(UK) part 2 written score[*]	1,266	3.68	6.85	−18.00	27.55
MRCP(UK) PACES score[*]	1,179	4.39	20.70	−64.00	42.00

*Exam score relative to the pass mark for that particular sitting.
CMT = core medical training; CPS = clinical problem solving test; MRCP(UK) = membership of Royal College of Physicians (UK); PACES = practical assessment of clinical examination skills; SJT = situational judgement test;

What is the incremental predictive validity of the MMTs over CMT selection methods in predicting MRCP(UK) examinations?

Hierarchical regression analyses explored the incremental validity of the MMTs over the CMT selection methods, with MRCP(UK) as the outcome. CMT shortlisting and interview scores were entered into the first step of the model, followed by the CPS and SJT in the second[b] (Table 3). CMT shortlisting and interview scores account for 16.7–30.0% of the variance in MRCP(UK) diploma results, while the CPS and SJT explain an additional 6.3–21.6%. Furthermore, Table 3 shows that for part 1 and 2 exams, the CMT selection methods are significant predictors, but after

the addition of the MMTs, their effects become less significant whereas the CPS is highly significant. Conversely, for PACES, the addition of MMTs reduces the significance of the CMT selection methods and the SJT becomes a significant predictor.

Discussion

In this study, we examined the longer-term predictive validity of the MMTs used in the GP selection process, along with the CMT selection methods, in predicting MRCP(UK) examination results.

All MMT and CMT selection methods were positively correlated with performance in all MRCP(UK) elements. In addition, the hierarchical regression analyses show that the MMTs add

Table 2. Correlations between selection assessments and MRCP(UK) examinations

		1	2	3	4	5	6
1	CMT shortlisting	–					
2	CMT interview	0.482*** n=1,974	–				
3	CPS	0.413*** n=2,275	0.559*** n=1,911	–			
4	SJT	0.357*** n=2,264	0.534*** n=1,902	0.524*** n=2,258	–		
5	MRCP(UK) part 1	0.486*** n=1,693	0.484*** n=1,379	0.692*** n=1,511	0.374*** n=1,502	–	
6	MRCP(UK) part 2 written	0.336*** n=1,266	0.386*** n=1,079	0.571*** n=1,142	0.369*** n=1,135	0.568*** n=1,227	–
7	MRCP(UK) PACES	0.345*** n=1,179	0.451*** n=1,021	0.379*** n=1,066	0.404*** n=1,061	0.319*** n=1,122	0.370*** n=1,150

n is smaller than total n for each correlation because of matching data.
**Correlation is significant at the 0.01 level
***Correlation is significant at the 0.001 level (2 tailed).
CMT = core medical training; CPS = clinical problem solving test; MRCP(UK) = membership of Royal College of Physicians (UK); PACES = practical assessment of clinical examination skills; SJT = situational judgement test.

Table 3. Multiple hierarchical regression examining predictors of performance in the MRCP(UK) diploma

	B	SE B	Lower bound 95% CI	Upper bound 95% CI	β
MRCP(UK) part 1 (n=1,325)					
Step 1 R^2=0.300, ΔF change=283.01**					
Constant	−38.917	1.766	−42.381	−35.453	-
CMT shortlisting	0.332	0.028	0.28	0.39	0.313***
CMT interview	0.515	0.043	0.43	0.60	0.320***
Step 2 R^2=0.516, ΔR^2=0.216, ΔF change=293.93**					
Constant	−62.569	1.929	−66.354	−58.784	-
CMT shortlisting	0.233	0.024	0.19	0.28	0.220***
CMT interview	0.122	0.042	0.04	0.20	0.076**
CPS	0.192	0.008	0.18	0.21	0.581***
SJT	−0.018	0.008	−0.03	−0.00	−0.056*
MRCP(UK) part 2 written (n=1,035)					
Step 1 R^2=0.167, ΔF change=103.54**					
Constant	−14.382	1.378	−17.087	−11.678	-
CMT shortlisting	0.090	0.019	0.05	0.13	0.152***
CMT interview	0.313	0.032	0.25	0.38	0.315***
Step 2 R^2=0.356, ΔR^2=0.189, ΔF change=151.18**					
Constant	−32.598	1.633	−35.803	−29.394	-
CMT shortlisting	0.045	0.017	0.01	0.08	0.076**
CMT interview	0.088	0.032	0.03	0.15	0.089**
CPS	0.097	0.006	0.09	0.11	0.462***
SJT	0.018	0.006	0.01	0.03	0.094**
PACES (n=979)					
Step 1 R^2=0.216, ΔF change=134.74**					
Constant	−62.040	4.291	−70.460	−53.619	-
CMT shortlisting	0.231	0.058	0.12	0.34	0.127***
CMT interview	1.213	0.098	1.02	1.41	0.393***
Step 2 R^2=0.279, ΔR^2=0.063, ΔF change=42.38**					
Constant	−93.871	5.555	−104.772	−82.970	-
CMT shortlisting	0.181	0.056	0.07	0.29	0.100**
CMT interview	0.760	0.107	0.55	0.97	0.247***
CPS	0.079	0.021	0.04	0.12	0.121***
SJT	0.132	0.019	0.10	0.17	0.224***

*Significant at the p<0.05 level
**Significant at the p<0.01 level
***Significant at the p<0.001 level;
SE = standard error; CI = confidence interval; ΔR^2 = R-squared change; CMT = core medical training; CPS = clinical problem solving test; MRCP(UK) = membership of Royal College of Physicians (UK); PACES = practical assessment of clinical examination skills; SJT = situational judgement test.

incremental validity (added value) to the CMT selection methods in predicting the subsequent performance in the MRCP(UK) part 2 and PACES. Although the SJT is significantly negatively associated with performance in the part one exam (B=−0.056, p<0.05), upon examining the 95% confidence interval for the beta coefficient (−0.02, −0.00), it is likely that no relationship exists between the SJT and the part 1 exam, or that it is in fact negligible. Therefore, these selection methods, including those originally designed for GP rather than CMT, have good

predictive validity in identifying applicants who perform well in MRCP(UK) examinations undertaken during CMT.

The results also show that the selection methods have differential prediction throughout the training pathway and for different types of criterion outcomes, with the CMT interview and SJT being stronger predictors of PACES, while the CPS was a stronger predictor of the MRCP(UK) written examinations (part 1 and part 2). This is to be expected as the SJT, CMT interview and PACES assess professional attributes centred on

communication with patients and professional integrity, whereas the CPS and MRCP(UK) written examinations both measure the application of clinical knowledge through similar formats. Taken together, our findings suggest that the assessment of non-academic attributes may become important later on in the training pathway, which is consistent with previous research.[5]

Implications

Our results confirm that the CMT selection process predicts subsequent exam performance in CMT, especially PACES. In addition, based on the available evidence, our results show that the MMTs may also be an appropriate assessment measure for shortlisting into CMT. The CPS offers potential value-add in predicting performance in part 1 and part 2 of the MRCP(UK) examination. Furthermore, the use of an SJT may offer an additional and standardised way to shortlist applicants prior to the interview phase. Thus, the MMTs could add value to the CMT selection process. Using the MMTs could be particularly useful since they are not only standardised, but more recently completed via computer-based methodology, which allows data to be collected instantly. That said, the implementation of any new selection method is not without controversy[6–8] and would require piloting for acceptability and deliverability. We would also recommend conducting job analysis research to ensure that any new machine marked tests were consistent with the CMT role. This approach has been taken in other, similar, healthcare roles.[9]

Limitations and recommendations for future research

One main limitation of this research that should be noted is that our use of examination outcomes may have limited the observed predictive validity of the selection methods. Examination scores at first attempt are generally more predictive of future performance; however, trainees are usually allowed multiple attempts at each exam. In terms of CMT/MMT selection scores, only scores relating to a successful offer were used (which may be less predictive than first attempt scores). The reason why these scores were used, though, is because candidates were allowed to apply to two CMT programmes across different regions, which could lead to variations in the selection process. Scores relating to a successful offer are therefore more relevant to validate since these are the ones used to rank candidates into their preferred placements.

We recommend that future research aims to establish whether the selection methods predict subsequent in-training performance, by examining the relationship between selection data and workplace-based assessment and/or supervisor ratings, for example. In particular, evidence has shown that SJTs are stronger predictors of performance when trainees enter higher clinical practice[10] and this would provide an opportunity to further understand this finding.

Conclusions

Overall, findings from this study offer good evidence of the validity of the CMT selection process. In particular, the results show that the CMT interview is a strong and valid predictor of performance in MRCP(UK) examinations. There is also some evidence that supports the value of the CPS and SJT in predicting performance in CMT selection and end of training, suggesting that they could be a practical methodology for adding further value to the CMT selection process. ∎

Conflicts of interest

Author contributions

SH and LB conceived of the original study and organised data collection. SL and FP contributed to the overall study design and methodology. SL analysed and interpreted the data, and wrote the paper along with FP. All authors commented on the original and final versions of the paper.

Acknowledgements

Notes

a In selection research, a correlation of 0.30–0.50 can be considered to be moderate, while above 0.50 is strong (particularly within the context of selection, where range restriction may occur at the lower end of score distributions).[11]

b A variance (R^2) value above 2% can be interpreted as a small effect size, 13% medium and 26% large.[12]

References

1 Carr A, Marvell J, Collins J. *Applying to specialty training: considering the competition*. London: BMJ Careers, 2013. http://careers.bmj.com/careers/advice/Applying_to_specialty_training%3A_considering_the_competition [Accessed 15 September 2016].

2 Patterson F, Ferguson E. Selection into medical education and training. In: Stanwick T (ed). *Understanding medical education: evidence, theory and practice*. Oxford: John Wiley & Sons, 2013.

3 Koczwara A, Ashworth V. Selection and assessment. In: Lewis R, Zibarras L (eds). *Work and occupational psychology: integrating theory and practice*. London: SAGE Publications, 2013:295–342.

4 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.

5 Lievens F. Adjusting medical school admission: assessing interpersonal skills using situational judgement tests. *Med Educ* 2013;47:182–9.

6 Harris BH, Walsh JL, Lammy S. UK medical selection: lottery or meritocracy? *Clin Med* 2015;15:40–6.

7 Harris BH, Walsh JL, Wilson DJ. The independent validation of the Foundation Programme application process: a closer look. *Clin Med* 2016;16:92–3.

8 Petty-Saphon K, Walker K, Patterson F, Ashworth V, Edwards H. Situational judgement tests reliably measure professional attributes important for clinical practice. *Adv Med Educ Pract* 2016;7:1–3.

9 Patterson F, Tavabie A, Denney M *et al*. A new competency model for general practice: implications for selection, training, and careers. *Br J Gen Pract* 2013;63:e331–8.

10 Lievens Fm Sackett PR. The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *J Appl Psychol* 2012;97:460–8.

11 Cohen J. *Statistical power analysis for the behavioral sciences*, 2nd edn. London: Routledge, 1998.

12 Field A. *Discovering Statistics Using SPSS*, 3rd edn. London: SAGE Publications, 2009.

Address for correspondence: Professor F Patterson, Work Psychology Group, 27 Brunel Parkway, Pride Park, Derby DE24 8HR, UK.
Email: f.patterson@workpsychologygroup.com