Article:
Tiffin, Paul Alexander, Morley, Emma, Paton, Lewis William et al. (2024) New evidence on the validity of the selection methods for recruitment to General Practice training:a cohort study. BJGP open. ISSN: 2398-3795

https://doi.org/10.3399/BJGPO.2023.0167

BJGP Open

RESEARCH

New evidence on the validity of the selection methods for recruitment to general practice training: a cohort study

Paul A Tiffin[1]*, Emma Morley[2], Lewis W Paton[1], Fiona Patterson[2]

[1]Health Professions Education Unit, Hull York Medical School, University of York, York, UK; [2]Work Psychology Group, Derby, UK

Abstract

Background: Selection into UK-based GP training has used the Multi-Specialty Recruitment Assessment (MSRA) and a face-to-face selection centre (SC). The MSRA comprises of a situational judgement test and clinical problem-solving test. The SC was suspended during the COVID-19 pandemic. Evidence is needed to guide national and international selection policy.

Aim: To evaluate the validity of GP training selection.

Design & setting: A retrospective cohort study using data from UK-based national recruitment to GP training, from 2015–2021.

Method: Data were available for 32 215 GP training applicants. The ability of scores from the specialty selection process to predict subsequent performance in the Clinical Skills Assessment (CSA) of the Membership of the Royal College of General Practitioners examination was modelled using path analysis. The effect sizes for sex, professional family background, and world region of qualification were estimated.

Results: All component scores of the selection process demonstrated statistically significant independent relationships with CSA performance ($P<0.001$), thus establishing their predictive validity. All were sensitive to demographic factors. The SC scores had the weakest relationship with future CSA performance. However, for candidates with MSRA scores below the lowest quartile, the relative contribution of the SC scores to predicting CSA performance was similar to that observed for MSRA components.

Conclusion: The MSRA has predictive validity in this context. Re-instituting an SC for those with relatively low MSRA scores should be considered. However, the relative costs and potential advantages and disadvantages should be carefully weighed.

***For correspondence:** paul.tiffin@york.ac.uk

Twitter: @ProfTiffin

Competing interest: See page 9

Received: 04 September 2023
Accepted: 24 November 2023
Published: 15 May 2024

Author Keywords: selection, training, predictive validity, general practice, primary health care, cohort studies

DOI:10.3399/BJGPO.2023.0167

How this fits in

The added value of a face-to-face selection centre (SC) for recruitment into GP training has been questioned. We found that performance on all components of GP training selection was independently related to eventual performance in the Clinical Skills Assessment (CSA). However, the relationship with the SC scores was relatively weak. Re-instituting SCs is unlikely to add substantial value in this context, although it could be considered for candidates with relatively low scores on the Multi-Specialty Recruitment Assessment (MSRA).

Introduction

There are well-recognised, long-term workforce shortages in GPs both in the UK and internationally.[1,2] The selection methods used to recruit applicants into GP training should accurately identify those

doctors likely to complete training and sustainably work effectively in primary care. Previous systematic reviews of international selection practices demonstrate the need to provide evidence of reliability, validity, and fairness.[3,4] There have been several previous studies evaluating data relating to the UK GP selection process[5,6] and this article presents new, large-scale evidence.

Since 2006, selection into UK GP training has been standardised and centralised through a national recruitment office.[7] Before the COVID-19 pandemic the process comprised of the following three stages:

- Stage 1: administrative, in which proof of eligibility and educational qualifications are checked.
- Stage 2: the MSRA is taken.
- Stage 3: SC, in which there is a face-to-face clinical examination and written test.

The MSRA consists of a situational judgement test (SJT; known as the professional dilemmas paper) and a clinical knowledge test (the Clinical Problem-Solving [CPS] paper). The MSRA aims to assess the clinical and interpersonal knowledge expected in a doctor completing their UK foundation years (FY) training. Training offers were determined by candidates' national ranking on their combined MSRA and SC scores. From 2016–2020 doctors scoring above an MSRA threshold (the 'bypass score') were exempt from the SC assessment. This score was 575 points, except in 2020, when it was 550. During the COVID-19 pandemic the SC stage was suspended and offer decisions were informed solely by the MSRA scores.

Once selected, doctors enter a 3-year programme of hospital and primary care posts. During training they must pass the Membership examination of the Royal College of General Practitioners (MRCGP). This includes written components (the Applied Knowledge Test; AKT) and Clinical Skills Assessment (CSA). Further details are included in Supplementary Information S1. Previously, it has been shown that performance on the MSRA validly predicts future achievement on the MRCGP.[8–10] Two separate studies also raised issues regarding the value of the SC,[5,6] since the scores accounted for only about 3% or 4% of additional variance once adjusted for MSRA performance. However, these studies treated the MSRA components as confounders rather than mediators of the relationship between SC and CSA performance. This introduced the risk of the so called 'table 2 fallacy'.[11] The authors also highlighted that the alternate forms reliability of the SC averaged only about 0.50 and also highlighted that abolishing the stage could save around £3 million over 3 years.[6]

Building on these prior investigations, the overall aims of this study were:

1. to evaluate the incremental validity (that is, the ability to predict the outcome, independent of the other measures) of the selection assessments in all applicants and in those with low MSRA scores; and
2. to evaluate the impact of key demographic characteristics (for example, sex and ethnicity).

The findings would guide policy on general practice and other specialty selection, both in the UK and elsewhere. For example, similar approaches are currently used for postgraduate selection in Australasia.[4]

Method

Ethics

The use of data from the UK Medical Education Database (UKMED) is not reliant on individual consent. However, any findings published from the UKMED must be presented in blunted form.[12] Thus, all frequencies are rounded to the nearest multiple of five.

Data processing and management

Data were available, via the UKMED, for 32 215 applicants to the GP training scheme during the period 2015–2021. The flow of study data is shown in ***Figure 1***.

Selection measures

The SJT comprises of 50 workplace-based scenarios linked to interpersonally oriented content domains including 'professional integrity', 'coping with pressure', and 'empathy and sensitivity'. Items use either a ranking response format or a multiple-choice format. The CPS paper has 97 questions in single best answer or enhanced matching format, assessing medical knowledge, presenting clinical

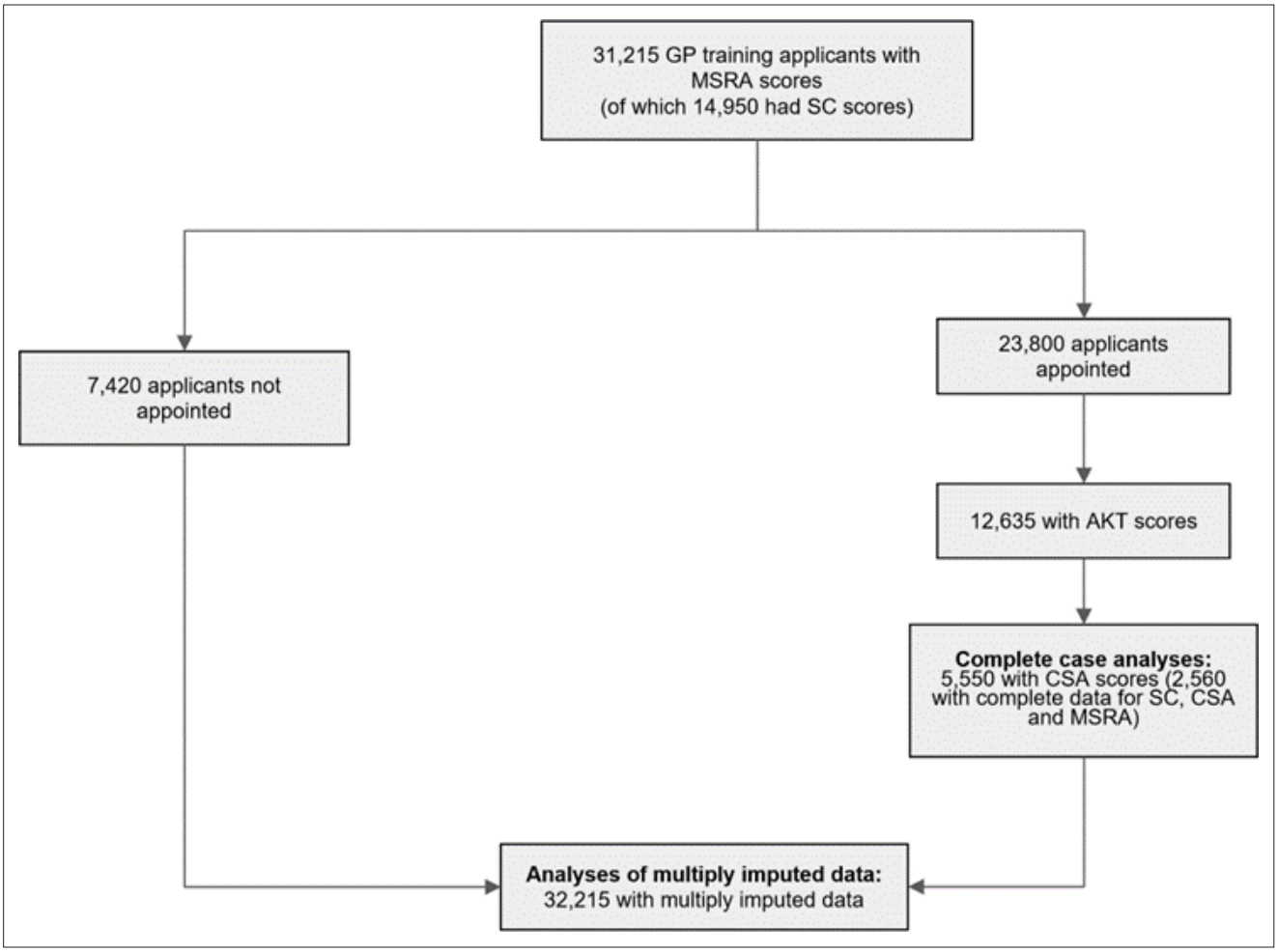

Figure 1 Flow of data through the study. Note that owing to blunting the numbers may not sum to the exact values. AKT = Applied Knowledge Test. CSA = Clinical Skills Assessment. MSRA = Multi-Specialty Recruitment Assessment. SC = selection centre.

scenarios for context. The scores are standardised across cohorts and summed to provide an overall MSRA score.

The SC is a set of face-to-face assessments, with three simulated consultation stations with actors and a written exercise. The latter was a short essay involving prioritising tasks in a clinical scenario.

To address the equivalence of scores on the MRCGP components over time, the marks 'relative to the pass' were used. 'Low MSRA' scores were classed as those below the bottom quartile for the study sample (466 marks). An explanation of why this definition was selected is provided in the Supplementary Information S4.

Missing data handling

The outcome (CSA score) was only observed in those appointed to the scheme and sitting the exam within the study timeframe. This 'range restriction' in personnel selection studies is a special case of missing data.[13] Consequently, missing values for both predictors and outcomes were imputed using chained equations. This is a valid approach to addressing this issue.[14–16] As a sensitivity analysis, imputed and non-imputed results were compared. Further details on the handling of missing data are available in the Supplementary Information S3.

Analysis approach

Univariable analyses were conducted. The potential impact of various educational and demographic characteristics (ethnicity, sex, and place of qualification) on the elements of the MSRA and the SC score were estimated via effect sizes. Either Cohen's *d* or Glass' Delta were calculated depending

Table 1 A breakdown of the demographic variables for those applicants to GP training with and without the primary outcome of interest (CSA score at first attempt)

Demographic variable	Applicants not entering scheme, *n* (%)	Entrants with at least one CSA attempt, *n* (%)	All applicants, *n* (%)	Missing values, *n* (%)
Male sex	3465/7420 (46.7)	2240/5550 (40.4)	12 800/31 215 (41.0)	0/31 215 (0.0)
Non-professional socioeconomic background	555/3145 (17.6)	685/3610 (19.0)	2995/15 490 (19.3)	15 725/31 215 (50.4)
BAME (UK graduates only)	1950/4485 (43.5)	1655/4440 (37.3)	7800/20 140 (38.7)	665/20 805 (3.2)
Place of qualifications				
UK	4655/7420 (62.7)	4585/5550 (82.6)	20 805/31 215 (66.7)	0/31 215 (0.0)
EEA	420/7420 (5.7)	190/5550 (3.4)	1520/31 215 (4.9)	0/31 215 (0.0)
IMG	2345/7420 (31.6)	775/5550 (14.0)	8895/31 215 (28.5)	0/31 215 (0.0)

BAME = Black, Asian and Minority Ethnic. CSA = Clinical Skills Assessment. EEA = European Economic Area. IMG = international medical graduate.

on equality of the intra-group variances. As some of the SC scores were imputed (owing to 'bypass' scores) the effect sizes were derived using the miesize package.[17] Ethnicity was self-reported and dichotomised into those identifying as White and those identifying as Non-White for the purposes of analysis. There was almost complete overlap between ethnicity and place of primary medical qualification (PMQ). Thus, ethnicity was only analysed in this respect for UK graduates. Also, when building the multivariable model, only the selection assessment scores were entered, as these were relevant to the selection decision.

Path analysis

Building multivariable models potentially gives rise to the 'table 2 fallacy'.[11] This assumes that all the predictor variables entered into the model are confounders, rather than mediators, moderators, or colliders ('reverse confounders'). This can be addressed by using structural equation modelling to model the underlying causal relationship between the variables. Thus, a path model was developed informed by prior research[18,19] (Supplementary Information S2 and Supplementary Figure S1). Stata/MP (version 17.0) and Mplus (version 8.8) were used to manage and analyse the data. All code is publicly available (https://github.com/pat512-star/P155).

Results

Table 1 summarises the demographic details for applicants and entrants. The educational and academic performance for applicants are included in Supplementary Table S1. From this point, all the results shown are derived from the imputed data (m = 10) unless otherwise indicated.

Univariable results

Univariable relationship between the measures

The correlations (Spearman's P values) between the key measures are shown in Supplementary Table S2. The results of the univariable analysis between the predictors and the CSA scores are shown in ***Table 2***. As can be seen, all three selection measure scores are significantly associated with predictive of CSA performance (P<0.001).

Sensitivity to demographic characteristics

The effect sizes for the three selection measures are shown in ***Figure 2*** (see also Supplementary Table S3). The largest effect sizes are observed for place of PMQ, with UK graduates scoring higher, on average, on all measures, compared with non-UK graduates. The SC scores are those most sensitive to sex (d = 0.42). However, the SC scores are less sensitive to socioeconomic status than the MSRA scores, although the confidence intervals overlap slightly in this respect. In contrast to the MSRA components, only a modest impact of ethnicity is observed for the SC scores. Note that the effect of ethnicity was only estimated in UK graduates.

Table 2 Results from the univariable and multivariable linear regression analyses predicting CSA performance from the scores from the three selection measures (CPS, SJT, and SC) on the multiply imputed study data (m = 10) for the whole sample and for different subgroups of applicants. The last three rows also report the results from analysis of the non-imputed data

Selection assessment	Coefficient (β)	*P* value	Lower 95% CI	Upper 95% CI	R^2 for the model[a]
Univariable results					
Clinical problem solving	0.17 (0.51)	<0.001	0.16	0.17	0.26
Situational judgement test	0.18 (0.56)	<0.001	0.17	0.19	0.31
Selection centre	0.91 (0.39)	<0.001	0.83	0.98	0.15
Multivariable results					
All applicants					
Clinical problem solving	0.09 (0.28)	<0.001	0.08	0.10	0.40
Situational judgement test	0.11 (0.34)	<0.001	0.10	0.12	
Selection centre	0.43 (0.18)	<0.001	0.33	0.53	
Applicants scoring below the first quartile on the MSRA					
Clinical problem solving	0.06 (0.15)	<0.001	0.05	0.07	0.12
Situational judgement test	0.08 (0.21)	<0.001	0.07	0.10	
Selection centre	0.42 (0.21)	<0.001	0.32	0.51	
Results from non-imputed data					
Clinical problem solving	0.10 (0.27)	<0.001	0.09	0.11	0.30
Situational judgement test	0.12 (0.32)	<0.001	0.10	0.13	
Selection centre	0.47 (0.17)	<0.001	0.37	0.56	

[a]This is the mean R^2 for the models derived from the imputed data. CPS = clinical problem-solving. CSA = Clinical Skills Assessment. SC = selection centre. SJT = situational judgement test.

Multivariable results

The multivariable results for the prediction of CSA performance from CPS, SJT, and SC are shown in ***Table 2***. Once the influence of SJT and CPS performance is controlled for, the SC ratings have only modest independent predictive ability in relation to the CSA score. Note, the overall predictive power of the SJT and CPS scores are weakened for those applicants with relatively low MSRA scores.

Path analysis results

A path analysis, based on the a priori hypothesised model, was estimated in the imputed datasets (m = 10). The initial model showed close to an acceptable fit to the data, according to the Comparative Fit Index (CFI) at 0.96. However, the Tucker–Lewis Index (TLI) was only around 0.89, where 0.90 is considered to indicate acceptable fit (see ***Figure 3***, Supplementary Figure S2, and Supplementary Table S4). Modification indices suggested considerable improvement in fit could be achieved by allowing the AKT score to be regressed on the SJT score. This suggested that the procedural knowledge tested by the SJT was also relevant to answering the AKT questions, and/or that performance on both assessments was at least partly determined by a shared ability. The resulting path model (***Figure 4A***) showed a good fit to the data, with the mean CFI being nearly 1.00 (0.995) and the TLI being 0.98. The fit indices for the models are shown in Supplementary Table S4. The path model for those with relatively low MSRA scores is also shown in ***Figure 4B***. The main difference to model A is the relatively reduced ability of the SJT scores to predict CSA performance. In contrast, the unique ability of the SC scores to predict CSA performance is consistent across models (β = 0.2). This infers around 4% of variance in the CSA scores is uniquely explained by the SC scores.

Results from the non-imputed data are shown in Supplementary Figures S3 and S4.

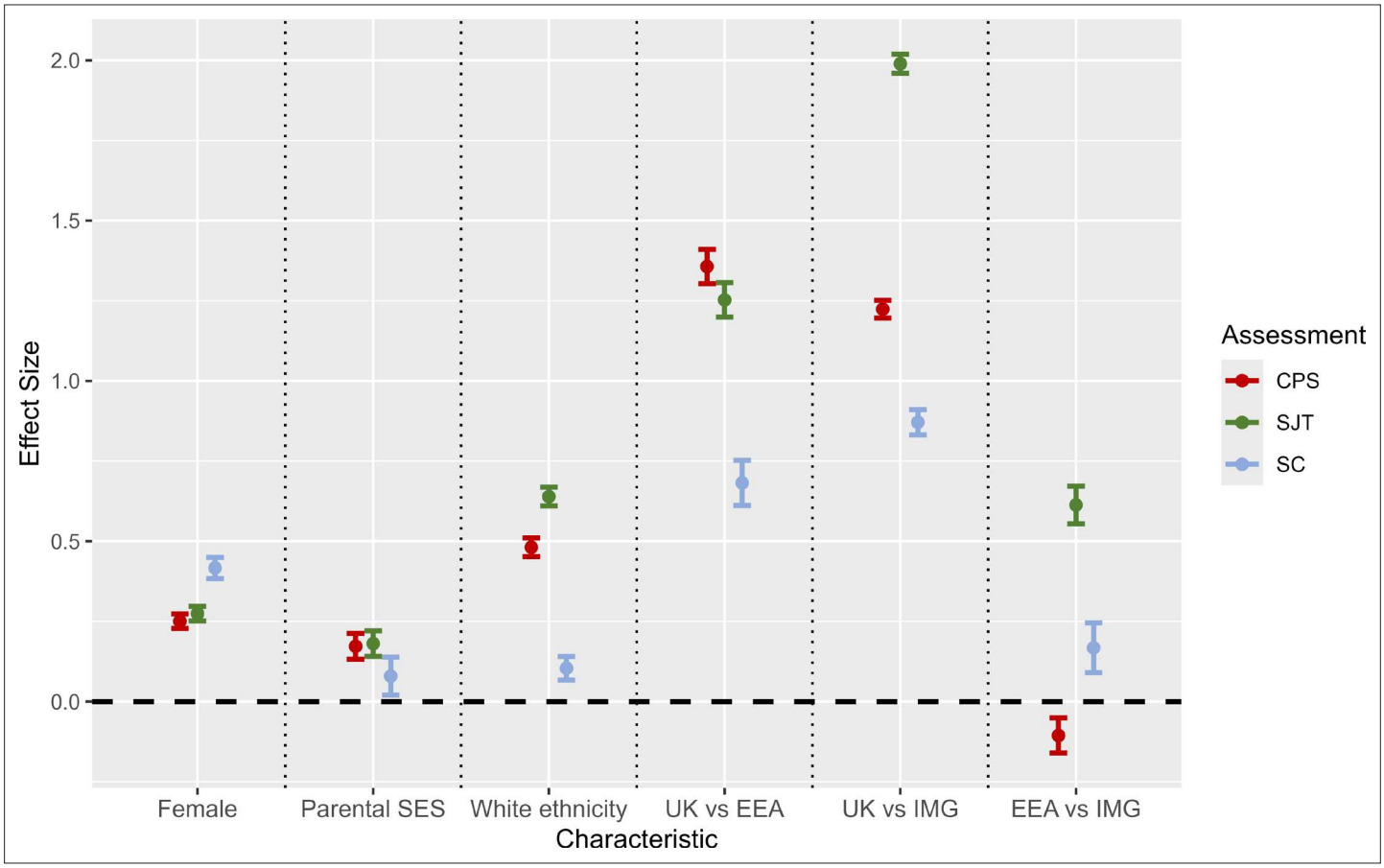

Figure 2 The effect size of key demographic characteristics on performance at the three selection assessments (clinical problem-solving [CPS] test, situational judgement test [SJT], and selection centre [SC]). EEA = European Economic Area. IMG = international medical graduate. SES = socioeconomic status according to occupational categorisation ('professional' versus 'non-professional').

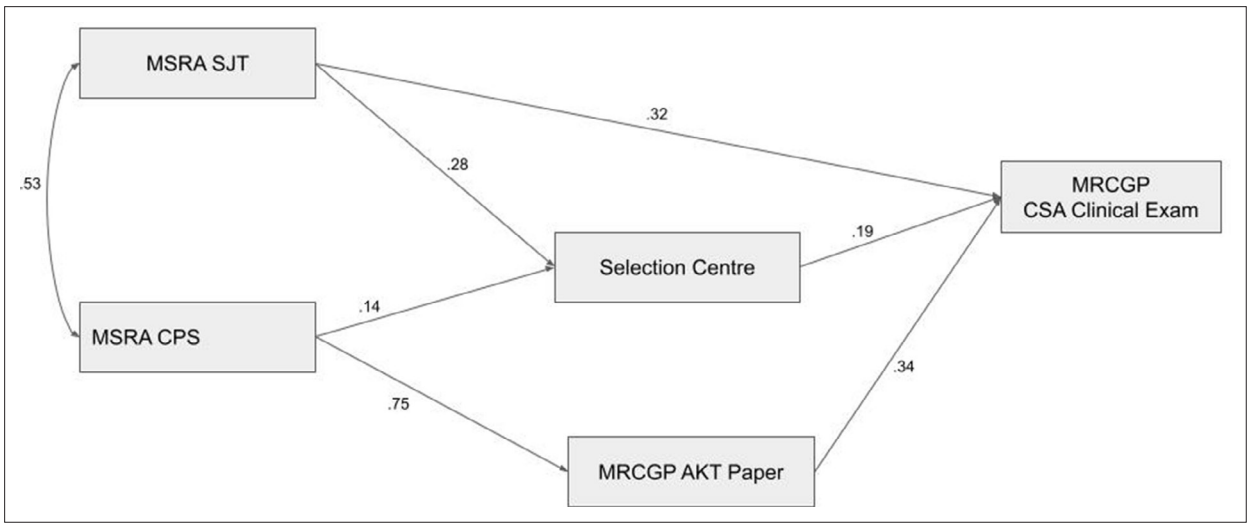

Figure 3 The a priori theoretical model for GP selection, testing the relationship between the predictors and outcome (CSA score) in the multiply imputed study data (*m* = 10) for all applicants (*n* = 31 215). AKT = Applied Knowledge Test. CPS = clinical problem-solving. CSA = Clinical Skills Assessment. MSRA = Multi-Specialty Recruitment Assessment. SJT = situational judgement test.

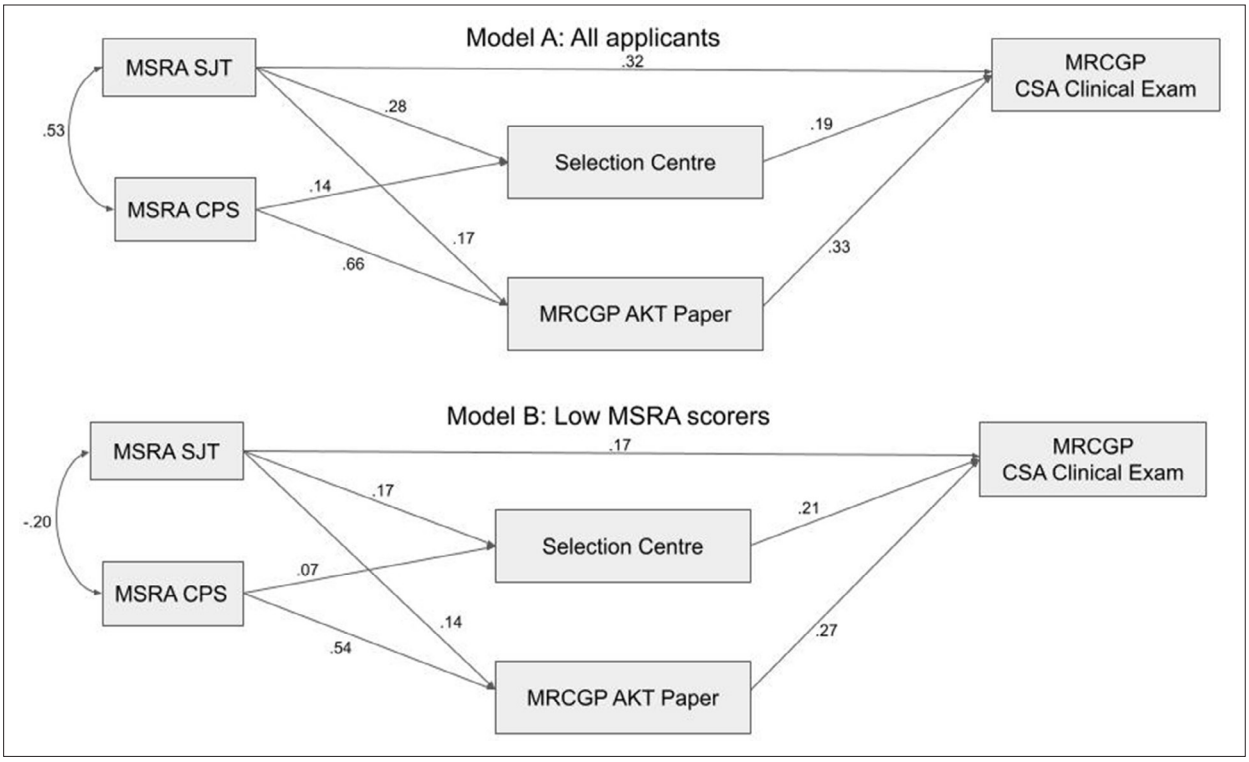

Figure 4 Path models modified according to modification indices, testing the relationship between the predictors and outcome (CSA score) in the multiply imputed study data (*m* = 10) for all applicants (Model A, *n* = 31 215) and for those scoring below the lowest quartile on the MSRA (Model B, *n* = 7795). AKT = Applied Knowledge Test. CPS = clinical problem-solving. CSA = Clinical Skills Assessment. MSRA = Multi-Specialty Recruitment Assessment. SJT = situational judgement test.

Discussion

Summary

In line with previous findings, we demonstrated that the MSRA scores predicted CSA performance. Also in line with prior findings, we observed the SC scores only incrementally predict an additional 3%–4% of variance in the CSA.[6] However, we also showed that, for applicants with low MSRA scores, this is also true for the SJT and CPS scores.

Our findings add to the GP-selection evidence in several important ways. First, our path analyses produced unbiased estimates of the unique contribution of each predictor. Second, we showed the relative reduced sensitivity of the SC scores, compared with the MSRA assessments, to key demographic characteristics. These have implications for equality and diversity in recruitment. Third, we estimated our models in applicants with relatively low MSRA scores to evaluate the relative value of SCs in this group of candidates.

Strengths and limitations

This was a relatively complete, national cohort of applicants. Nevertheless, there was the obvious challenge of not being able to observe the outcome of interest in those who had not been appointed or undertaken the CSA within the study timeframe. However, this issue was addressed using multiple imputation. Our imputed and non-imputed results did not meaningfully differ, providing evidence for the validity of this approach. Ideally our validity-related outcome would have been aspects of actual workplace behaviour. However, as these were not available, high-fidelity clinical simulation examinations may be the best available proxy for this. In this respect, US-based research evidence suggests that physician performance in postgraduate cardiology examinations predicts clinical outcomes in their patients experiencing myocardial infarction.[20]

Comparison with existing literature

These results also have important general implications for the design of selection processes within postgraduate medicine and add to the international evidence base examining the validity of differing approaches.[4] These relate to how well the constructs evaluated by selection assessments map to the validity criterion chosen. For example, in contrast, a similar evaluation of selection into psychiatry training noted that SC scores were moderately independently correlated with performance in the clinical membership examination.[19] This may be because the SC used in psychiatry more closely resembles the subsequent clinical examination. Certainly, our findings are in keeping with the existing evidence for the use of selection approaches in postgraduate medical selection.[4] That is, there is relatively strong evidence for the validity of scores derived from SJTs, CPS tests, and multiple mini-interviews. There is less consistent evidence for the validity of using curriculum vitae, references, and personal statements.

The sensitivity of the differing selection components to demographic factors have implications for equality and diversity. Specifically, in this context, the SJT was more sensitive to ethnicity and PMQ region than the other measures. Depending on their characteristics, such as the complexity of language or contextualisation of content, selection SJTs can be differentially sensitive to demographic factors.[21] Thus, this issue is worth further researching. Differential attainment in clinical educational test scores between UK and non-UK medical graduates has been observed across numerous medical specialties.[22] The drivers behind such phenomena are acknowledged to be complex and may include factors such as language fluency and cultural factors.[18,23] In contrast, SC performance was less sensitive to ethnicity than the MSRA in this context. However, the relative insensitivity of the SC scores to demographics may be, at least partly, an artefact of its relatively low reliability.

Implications for research and practice

Our findings support the continued use of the MSRA in this context. Moreover, they suggest that the re-introduction of some form of SC should be reconsidered for a smaller number of 'borderline' applicants, with relatively low scores on the MSRA. It is also possible that SCs could be made more reliable; for example, by reconfiguring the time allocated to create an increased number of shorter stations that could be delivered online, which would also significantly reduce costs.[24] With the exception of sex, the SC scores seemed less sensitive to demographic factors, including ethnicity and world region of qualification. This implies placing some weight on an SC score, compared with the MSRA performance, could widen access to GP training. This should be considered and further explored. It may also be that candidates who perform poorly at SCs may benefit from early additional support and subsequently succeed at postgraduate training.[25]

The shortage of GPs in the workforce also needs to be considered. A previous study simulating changes to the GP selection system highlighted that completely removing MSRA cut-off scores would increase the number of trainees more than changing the selection process in other ways. However, it would also likely significantly increase the number of doctors failing to complete their GP training.[26] There may also be patient safety and care-quality issues if selection procedures are not sufficiently robust. Thus, the issue of an MSRA cut-off ('bypass') score is complex. Further research could investigate how selection processes relate to other important outcomes, such as retention in the primary care workforce. Further modelling, taking a 'Pareto-optimal' perspective,[27] could also be helpful in locating the optimum trade-off between educational performance and numbers of GP trainees recruited. Ideally, future evaluations should include cost-benefit analyses. These will inevitably be complex. They will have to account for both relatively direct costs (for example, temporary staff for vacant posts) and indirect costs (for example, those related to the risk of complaints and compensation relating to poor clinical practice). Such economic modelling will also take place in the context of a shifting workforce landscape with unpredictable trends relating to staff retention and medical migration, both to and from the UK.

The selection process is valid, in that performance on each component independently predicts future CSA performance. The MSRA scores are more predictive than those for the SC for CSA performance, but not for those with low MSRA scores. Thus, the use of an online SC for a relatively small number of 'borderline' candidates should be considered. This could address diversity issues and widening access to UK GP training. It may also optimise the absolute numbers of GP trainees. However, the potential costs and disadvantages should be weighed when making this decision.

Funding

This study was partly funded by Health Education England (HEE). Any views expressed in this report are those of the authors and do not necessarily reflect those of HEE or NHS England.

Ethical approval

As the study used routinely collected, de-identified data, ethical approval was not required. This was confirmed in writing by the Chair of the University of York Health Sciences Ethics Committee. Also, as the data were held within UKMED, individual consent was not required. This is because the data are used for the statutory functions of the GMC, via powers granted by the Medical Act (1983). Analyses were conducted in a 'safe-haven' where only de-identified data are shared with researchers and only summary reports, not individual data, can be extracted. Data access is granted via project approval by the UKMED research subgroup. Any published findings are presented in blunted form in accordance with HESA statistical disclosure controls, with all frequencies rounded to the nearest multiple of 5.

Provenance

Freely submitted; externally peer reviewed.

Acknowledgements

The authors would like to thank Daniel Smith (at the General Medical Council [GMC]) for their helpful assistance in accessing the data used in this study, held in the UK Medical Education Database (UKMED). We are also grateful to the members of the UKMED Research Subgroup for their helpful comments and advice on an earlier version of this report. The source of the data used in this project is the UKMED (project number: P155), with the extract generated on 29 September 2022, reissued 2 March 2023. Approved for publication on 15 November 2023. We are grateful to UKMED for the use of these data. However, UKMED bears no responsibility for their analysis or interpretation. The data include information derived from that collected by the Higher Education Statistics Agency Limited (HESA) and provided to the GMC (HESA Data). Source: HESA Student Record 2007/2008 and 2008/2009 © HESA. HESA makes no warranty as to the accuracy of the HESA Data and cannot accept responsibility for any inferences or conclusions derived by third parties from data or other information supplied by it. The authors also wish to thank Máire Kerrin (Work Psychology Group) for their useful feedback on an earlier version of this report.

Competing interests

Fiona Patterson and Emma Morley are employed by Work Psychology Group, who provide consulting advice on selection methods (including knowledge tests, situational judgement tests, interviews, and selection centres) to a range of organisations internationally. Specifically, they are also commissioned by NHS England to design the Multi-Specialty Recruitment Assessment. The other authors have declared no competing interests.

References

1. Centre for Workforce Intelligence. *In-depth review of the general practitioner workforce: final report*. 2014. https://assets.publishing.service.gov.uk/media/5a7ff981ed915d74e33f7b37/CfWI_GP_in-depth_review.pdf (accessed 26 Apr 2024).
2. British Medical Association. Pressures in general practice data analysis. 2024. https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/pressures/pressures-in-general-practice-data-analysis (accessed 26 Apr 2024).
3. Patterson F, Knight A, Dowell J, *et al*. How effective are selection methods in medical education? A systematic review. *Med Educ* 2016; **50**(1): 36–60. DOI: https://doi.org/10.1111/medu.12817
4. Roberts C, Khanna P, Rigby L, *et al*. Utility of selection methods for specialist medical training: A BEME (best evidence medical education) systematic review: BEME guide No.45. *Med Teach* 2018; **40**(1): 3–19. DOI: https://doi.org/10.1080/0142159X.2017.1367375
5. Botan V, Williams N, Law GR, Siriwardena AN. *How is performance at selection to general practice related to performance at the endpoint of GP training?* 2022. https://repository.lincoln.ac.uk/articles/report/How_is_performance_at_selection_to_general_practice_related_to_performance_at_the_endpoint_of_GP_training_/24396313 (accessed 8 May 2024).
6. Davison I, McManus C, Taylor C. *Evaluation of GP specialty selection*. 2016. https://www.ucl.ac.uk/medical-education/sites/medical-education/files/GPspecSelReport.pdf (accessed 26 Apr 2024).

7. NHS England. *General Practice (GP) recruitment hub*. 2023. https://medical.hee.nhs.uk/medical-training-recruitment/medical-specialty-training/general-practice-gp (accessed 26 Apr 2024).
8. Patterson F, Lievens F, Kerrin M, *et al*. The predictive validity of selection for entry into postgraduate training in general practice: evidence from three longitudinal studies. *Br J Gen Pract* 2013; **63**(616): e734–e741. DOI: https://doi.org/10.3399/bjgp13X674413
9. Patterson F, Kerrin M, Baron H, Lopes S. *Exploring the relationship between general practice selection scores and MRCGP examination performance. September 2015 final report.* 2015. https://www.gmc-uk.org/-/media/documents/Exploring_the_Relationship_between_Recruitment_Scores_and_MRCGP_Examination_Performance_v5.4.pdf_63533914.pdf (accessed 26 Apr 2024).
10. Siriwardena AN, Botan V, Williams N, *et al*. Performance of ethnic minority versus white doctors in the MRCGP assessment 2016–2021: a cross-sectional study. *Br J Gen Pract* 2023; **73**(729): e284–e293. DOI: https://doi.org/10.3399/BJGP.2022.0474
11. Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am J Epidemiol* 2013; **177**(4): 292–298. DOI: https://doi.org/10.1093/aje/kws412
12. Higher Education Statistics Agency. Rounding and suppression to anonymise statistics. https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics (accessed 26 Apr 2024).
13. Dahlke JA, Wiernik BM. Not restricted to selection research: accounting for indirect range restriction in organizational research. *Organ Res Methods* 2020; **23**(4): 717–749. DOI: https://doi.org/10.1177/1094428119859398
14. Zimmermann S, Klusmann D, Hampe W. Correcting the predictive validity of a selection test for the effect of indirect range restriction. *BMC Med Educ* 2017; **17**(1): 246. DOI: https://doi.org/10.1186/s12909-017-1070-5
15. Mwandigha LM. *Evaluating and extending statistical methods for estimating the construct-level predictive validity of selection tests.* 2017. https://etheses.whiterose.ac.uk/21267/1/Lazaro_Mwakesi_Mwandigha_PhD_thesis.pdf (accessed 26 Apr 2024).
16. van Ginkel JR, Linting M, Rippe RCA, van der Voort A. Rebutting existing misconceptions about multiple imputation as a method for handling missing data. *J Pers Assess* 2020; **102**(3): 297–308. DOI: https://doi.org/10.1080/00223891.2018.1530680
17. Tiffin PA. MIESIZE: Stata module to estimate effect sizes from multiply imputed data. 2024. https://econpapers.repec.org/software/bocbocode/s459181.htm (accessed 26 Apr 2024).
18. Patterson F, Tiffin PA, Lopes S, Zibarras L. Unpacking the dark variance of differential attainment on examinations in overseas graduates. *Med Educ* 2018; **52**(7): 736–746. DOI: https://doi.org/10.1111/medu.13605
19. Tiffin PA, Morley E, Paton LW, *et al*. The validity of the selection methods for recruitment to UK core psychiatry training: cohort study. *BJPsych Bull* 2024; 1–10. DOI: https://doi.org/10.1192/bjb.2024.9
20. Norcini JJ, Lipner RS, Kimball HR. Certifying examination performance and patient outcomes following acute myocardial infarction. *Med Educ* 2002; **36**(9): 853–859. DOI: https://doi.org/10.1046/j.1365-2923.2002.01293.x
21. Lievens F, Patterson F, Corstjens J, *et al*. Widening access in selection using situational judgement tests: evidence from the UKCAT. *Med Educ* 2016; **50**(6): 624–636. DOI: https://doi.org/10.1111/medu.13060
22. Tiffin PA, Illing J, Kasim AS, McLachlan JC. Annual Review of Competence Progression (ARCP) performance of doctors who passed Professional and Linguistic Assessments Board (PLAB) tests compared with UK medical graduates: national data linkage study. *BMJ* 2014; **348**: g2622. DOI: https://doi.org/10.1136/bmj.g2622
23. Roberts C, Atkins S, Hawthorne K. *Performance features in clinical skills assessment: linguistic and cultural factors in the Membership of the Royal College of General Practitioners examination*. London: King's College London; 2014.
24. Callwood A, Gillam L, Christidis A, *et al*. Feasibility of an automated interview grounded in multiple mini interview (MMI) methodology for selection into the health professions: an international multimethod evaluation. *BMJ Open* 2022; **12**(2): e050394. DOI: https://doi.org/10.1136/bmjopen-2021-050394
25. Brown J, Jenkins L, Sandars J, *et al*. *Evaluation of the impact of the Royal College of Psychiatrists Clinical Assessment of Skills and Competencies masterclass on reducing the attainment gap*. 2023. https://www.gmc-uk.org/-/media/documents/the-casc-masterclass---interim-evaluation-2023_pdf-101480061.pdf (accessed 26 Apr 2024).
26. Taylor C, McManus IC, Davison I. Would changing the selection process for GP trainees stem the workforce crisis? A cohort study using multiple-imputation and simulation. *BMC Med Educ* 2018; **18**(1): 81. DOI: https://doi.org/10.1186/s12909-018-1160-z
27. Lievens F, Sackett PR, De Corte W. Weighting admission scores to balance predictiveness-diversity: the Pareto-optimization approach. *Med Educ* 2022; **56**(2): 151–158. DOI: https://doi.org/10.1111/medu.14606