Unpacking the 'Dark Variance' of Differential Attainment: Predictors of Performance in a Clinical Postgraduate Exam in International Medical Graduates

1* Professor Fiona Patterson, University of Cambridge and Work Psychology Group, UK.
2 Dr Paul A. Tiffin, Health Sciences, University of York, UK.
3 Safiatu Lopes, Work Psychology Group, UK.
4 Dr. Lara Zibarras, City, University of London, UK and Work Psychology Group, UK.

Authorship & Contributorship

All authors have made a substantial contribution to the design, the intellectual content, the data interpretation and analysis, and writing of the paper. FP and PAT conceived of the overall study and analysis strategies, and led the interpretation of results. PAT also contributed to the statistical analyses. SL assisted in data collection, analysis, interpretation of results and in drafting the paper. All authors helped draft several versions of the paper and provided comments. All authors commented on the final version of the paper.

Acknowledgements

Many thanks go to Kirsty White, Daniel Smith and Kerrin Clapton at the GMC and to the RCGP for assistance with facilitating access to the data used in this study. Thanks also goes to Helen Baron who was involved in the original study funded by the GMC.

Ethical Approval

The analysis was performed on de-identified, routinely arising data. Thus an external ethical review and opinion was not required. This was confirmed in writing by the Chair of the Research Ethics committee for the University of York, Department of Health Sciences.

Competing Interests

FP, SL and LZ provide advice to and receive funding from Health Education England in designing the GP selection methodology (including the clinical knowledge test and situational judgement test) through the Work Psychology Group. PAT has previously received research funding from the Economic and Social Research Council (ESRC), the Engineering and Physical Sciences Council (EPSRC), the Department of Health for England, the UKCAT Board, and the General Medical Council (GMC). In addition, PAT has previously performed consultancy work on behalf of his employing University for the UKCAT Board and Work Psychology Group and has received travel and subsistence expenses for attendance at the UKCAT Research Group.

Guarantor

FP is guarantor for the paper.

Source of Funding

This project was funded as a research grant by the GMC. PAT is currently supported in his research by a National Institute for Healthcare Research (NIHR) Career Development Fellowship. This paper presents independent research part-funded by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. PAT is also lead for the DREAMS Network, an international collaboration on selection into the professions, of which FP is also a member, which is funded by a Worldwide University Network (WUN) Research Development Fund award.

Abstract

Context

Differential performance at postgraduate exams between home medical graduates and those who qualified outside their country of practise is well recognised. This difference is especially marked in the practical component of the UK Membership of the Royal College of General Practitioners (MRCGP) exam. The potential causes of such disparities are not well understood.

Methods

Data were available for 1874 international medical graduates who applied for UK GP speciality training 2008-2012. The primary outcome was performance at the Clinical Skills Assessment (CSA) OSCE component of the MRCGP. The main predictors were performance on the Situational Judgment Test (SJT) and the Clinical Problem Solving Test (CPST- a test of applied clinical knowledge), used in the selection for GP training. Data relating to the demographic characteristics and English language fluency were also available. To understand better the relationship between the predictors, the selection measures, and the outcome, a series of univariable and multivariable models were developed and tested, concluding with a structural equation model to explore causality.

Results

The CSA rating was more strongly predicted by SJT scores (standardised beta 0.26) than by performance on the CPST (standardised beta 0.17). There was a relationship between English language fluency and CSA score that was mainly mediated via SJT performance.

Conclusions

These findings demonstrate that performance on an SJT predicts performance at a high fidelity clinical simulation (the CSA) in international medical graduates. Whilst the constructs tested by SJTs are debated, and are likely to vary across settings, culturally appropriate knowledge of interpersonal competence is likely to be evaluated. Improving the confidence of doctors in this area through targeted educational interventions, rather than focussing on increased clinical knowledge, is likely to be more effective at reducing disparities observed in postgraduate exam performance. Thus there are important implications for the design of speciality selection and licensing assessments globally.

Unpacking the 'Dark Variance' of Differential Attainment: Predictors of Performance in a Clinical Postgraduate Exam in International Medical Graduates

INTRODUCTION

Doctors wishing to practise as general practitioners (family physicians) in the UK are required to pass the Membership of the Royal College of General Practitioners (MRCGP) licensing examination.(1) Most trainees pass this exam at first attempt with a minority of postgraduate trainees struggling to meet the criteria for independent practice.(2) However, international medical graduates (i.e. those who obtained their primary medical qualification from outside the European Economic Area) have a substantially lower pass rate compared to UK graduates.(3) Differential attainment is particularly marked for the Clinical Skills Assessment (CSA) OSCE component of the MRCGP. The CSA involves the observation and rating of candidates interacting with clinical scenarios. Esmail and Roberts reported that Black and Minority Ethnic (BME) international medical graduates were 14.7 times more likely to fail the CSA at first attempt compared to white UK graduates, even after controlling for the potential influence of age, sex and performance in the written, knowledge-based component of the MRCGP- the Applied Knowledge Test (AKT).(4) These differential pass rates for the CSA were subsequently debated in the High Court when the British Association of Physicians of Indian Origin (BAPIO) took legal action (ultimately unsuccessfully) against the GMC and the Royal College of General Practitioners.(5)

These observations are not restricted to the specialty of general practice; meta-analytic studies consistently report differences in candidate performance relating to ethnicity across a variety of medical education settings internationally.(6,7) Whilst BME candidates, on average, tend to perform less well than comparable White candidates in medical educational academic assessments the effect is confounded by place of medical qualification.(2) It is also notable that the most marked differential outcomes between home and international medical graduates are for specialisms where patient-doctor communication was a key focus, such as general practice and psychiatry.(8)

Despite these consistent findings there is no firm causal explanation for the observed differences in attainment(9) which ultimately limits the scope for high quality educational interventions. Indeed, it has been highlighted that, generally, around a third of variation in medical academic performance is unexplained by prior educational attainment. An

astrophysical analagy has been used for these poorly characterised, presumably non-academic attributes, that may explain the '*dark variance'* of medical selection.(10) English language fluency and clinical knowledge and skills are two factors that may influence the discrepancy between international medical graduates and UK graduates.(3,4,8) Specifically, international medical graduates may struggle with the nuances of the English language and cultural issues.(11,12)

Previous research exploring differential performance has mainly focused on the final stage of assessment of competence for GPs; that is, the MRCGP. In contrast we examine differential attainment at the point of selecting doctors into GP training. Currently the cost of GP training is estimated at £485,390 per trainee.(13) Thus, more effective staff selection will ultimately lower the costs (both direct and indirect) when doctors fail to progress in training. Since 2009 there has been a decline in both training and practicing GPs.(14) Given these trends it is important to maximise the number of trainees who qualify. If trainees who may struggle to complete the programme are identified early, appropriately targeted support could promote a higher qualification rate, reducing overall training costs. This was recently estimated as £64,395 for every 6 month period of extension. This includes costs for the extension, remediation, administration and loss of healthcare service which the trainee would have provided.(15)

The current GP selection process is a three stage, competency-based selection process (see Figure 1) based on a job analysis (16,17) which predicts in-training performance.(18,19). Applicants are scored against attributes considered important in relation to GP training (e.g. *empathy*, *communication*, *clinical expertise*). An aggregated individual summary score is produced. In Stage 1 applicants are long-listed based on eligibility criteria. International medical graduates, from Countries where English is not widely spoken as a first national language, must also achieve minimum scores on the International English Language Testing System (IELTS) test and the Professional and Linguistic Assessments Board (PLAB) (if they are not from the European Economic Area).(20)

INSERT FIGURE 1 AROUND HERE

In Stage 2 candidates complete both a computer-based Clinical Problem Solving Test (CPST) and a Situational Judgement Test (SJT). The CPST is a test of clinical knowledge and measures the ability to apply this when making clinical decisions. The SJT presents, in written format, a series of scenarios. Candidates are asked to rank a list of possible behavioural responses to the scenarios according to their perceived effectiveness or

appropriateness.(21). The SJT content is designed around a number of domains (*empathy*, *professional integrity* and *coping with pressure*) identified as critical to competent performance in the role.(18,22) Thus, the SJT could be considered as providing a metric of a person's knowledge of interpersonal competence (23) within a primary care health service context.

Finally, in Stage 3, successful candidates take part in a selection centre based on a multi-trait, multi-method assessment approach. The selection centre tests aptitude for training in GP, following which allocation to local education and training boards (LETB) takes place based on trainees' preferences and their aggregated score across the three stages. Once candidates are accepted onto GP specialty training their progress performance is measured through an integrated MRCGP assessment system (see Figure 1), involving the AKT and CSA exams as well as Workplace Based Assessments (WPBAs).

It could be argued that clinical knowledge (as measured by the CPST or AKT) is relatively comparable across languages because the content is objective. Conversely, the practical exams, such as the CSA, may require a degree of interpersonal and cultural competence.(4,21) Likewise, performing well on the SJT is likely to require an ability to recognise both professionally and culturally appropriate (and inappropriate) behaviours and make judgments about them. In this study, in the absence of data related to patient outcomes or experience, we used the CSA as a 'high-fidelity simulation' proxy of clinical behaviour in practice. We hypothesised that, in a group of international medical graduates, language performance on the SJT would have a relatively larger influence on performance at the CSA compared to tests of clinical knowledge. In turn, we anticipated that English Language fluency, as evaluated by the IELTS, would act as something of a filter, having an impact on both the predictors and outcome of interest, in line with knowledge acquisition models (24). Thus, we proposed a model that hypothesised a relationship between language fluency (an eligibility criteria), the two stage 2 selection measures (CPST and SJT) and the CSA rating (Figure 2).

INSERT FIGURE 2 AROUND HERE

Thus, the aim of this retrospective observational cohort study was to understand the determinants of (simulated) clinical performance in candidates applying for GP training in the UK, in relation to the selection measures. Our findings would have implications for how to effectively support practitioners working outside of their country of qualification, as well as to

regulatory policy. As Western countries, such as the UK, will continue to rely on international graduates to deliver health services tackling such disparities is important to the wellbeing of overseas doctors as well as patient care.(25)

METHOD

Data

Data were provided by the General Medical Council (GMC), the UK medical regulator, supplied in a de-identified form to the research team. (26) Demographic data (place of medical qualification, sex, ethnicity and age) were available. Selection data for medical trainees who applied to GP specialty training between 2008 and 2012 were provided to the GMC by the GP National Recruitment Office in the form of standardised CPST and SJT scores. In the analyses, the scores from the successful (last) attempt at CPST and SJT resulting in a training place offer were used since this is what selectors base their decision on. Moreover, in order to evaluate how the achieved level of applied knowledge and performance on the SJT influenced CSA rating it was considered that the most recent scores would reflect this level of attainment. Overall and component IELTS band scores (*listening*, *speaking*, *writing* and *reading*) were provided by the GMC and matched for trainees who had applied for GP training. The IELTS overall scores are derived by averaging the subtest bands and rounding up to the nearest half integer.(27)

MRCGP performance data for the years 2008 to midway through 2013 were provided by the Royal College of GPs (RCGP). The CSA scores were also available for only a proportion of the candidates, as some may have failed the GP selection process, declined an offer, received a placement or either left/paused training or were still due to sit the examination. This study used scores from the first CSA attempt, expressed as relative to the pass mark for that particular sitting as the primary outcome. Trainees are allowed up to a maximum of four sittings of the CSA, so the first attempt may not be a true reflection of typical performance. However, other approaches (e.g. an average of examination performance) are equally problematic since subsequent scores are often higher and those who failed would have more than one score. Therefore, in this instance we considered it is more meaningful to use first attempt data.(22) We also had access to the AKT results for GP training scheme applicants. However, these scores were only included in the univariable not the multivariable models. This was because AKT was not used as primary selection measure for entry to GP training, as it is taken subsequently. Moreover, both the CPST and AKT evaluate applied clinical knowledge, albeit calibrated to different stages of training. This was evidenced by the relatively high correlation between the two scores at first attempt (r=0.61). Note that this

correlation was slightly higher than that observed in the scores at pass (subsequently used) as the latter were left (lower) censored at the pass mark. In this study, only candidates who had IELTS scores were included (predominantly international medical graduates, with a small number from the European Economic Area).

Data analysis

Scores for the CPST, SJT, and CSA were normally distributed. The IELTS scores were positively skewed and so Spearman's rho was employed as a non-parametric index of correlation. The conceptual model (Figure 2) was tested with a Structural Equation Modelling (SEM) framework. Within the SEM English fluency was modelled as a latent variable with the four IELTS subset bands serving as ordinal indicators. The SJT, CPST and CSA scores were defined as endogenous (dependent) observed (manifest) variables. The SEM was estimated using robust weighted least squares estimation (i.e. WLSMV), as a robust estimator which does not assume normally distributed data and also accommodates the ordinal nature of the IELTS band scores.(28) Model fit was evaluated using the Comparative Fit Index (CFI) (29), the Tucker-Lewis Index (TLI)(30) and the root mean square error of approximation (RMSEA)(31,32). For the SEM standard errors and confidence intervals were derived via a bootstrapping process, with a 1000 replications and correction for bias. (33)

STATA 14.2 MP(34) was used to conduct the correlation, univariable and multivariable regression analyses. MPlus v8 was used for the path analyses.(35)

RESULTS

For doctors with data on selection measures scores available, in terms of ethnicity, 2149 (57.3%) classified themselves as 'Asian', 385 (10.3%) as 'Black', 178 (4.7%) as 'White', 198 (5.3%) as 'Mixed & Other', while reported ethnicity was missing in 840 (22.4%) cases. The average age for the sample was 34 years old (SD = 4.60 years). In the sample 1,889 (50.4%) were male. Table 1 displays the descriptive statistics for the sample who had at least some data on IELTS or selection scores (n=3750). Table 1 also shows the degree of correlation between the background, predictor and outcome variables. As can be seen, there were at least moderate correlations observed between IELTS reading band, AKT, CPST and SJT scores (r ranging from 0.18 (CPST) to 0.29 (AKT)). It should be noted that, for consistency, the correlations were performed on the most recent (pass) sittings of the CPST, SJT and AKT and the first attempt at CSA. Although this would have led to attenuation in the observed correlations (due to lower (left) censorship at the pass marks for the predictors) these were the scores used in the later univariable and multivariable models. Likewise the

limited range and categories for the IELTS bands would be expected to result in relatively low observed correlations between the subtest scores.

INSERT TABLE 1 AROUND HERE

The results from univariable regression analyses, predicting CSA scores from the background, demographic and selection measures are presented in Table 2. As can be seen, being female and younger were all significantly associated with higher CSA scores. Likewise, better performance at IELTS (especially *reading* and *listening*), AKT, the SJT and CPST was also associated with, on average, higher scores at CSA. The effect of gender on the selection measures was also evaluated. In the study sample there was no sex difference observed for the SJT compared to males (p=0.7). There was a slight, non-statistically significant trend for women to score more highly on the CPST (mean score 238.26 vs 236.48, p=0.1). Males also reported, on average significantly lower IELTS overall scores (mean score 7.38 vs 7.47, p<0.001).

INSERT TABLE 2 AROUND HERE

In order to assess the independent effects of the predictor variables they were entered into a multivariable linear regression model. (36). Note that the IELTS overall score was included but that the subtest scores were not entered individually as it would not be possible to meaningfully interpret each subtest band independent of the other three. However, the relative strength of the univariable relationships between the IELTS subtest scores and the CSA ratings can be viewed in Tables 1 and 2. As can be seen in Table 3, all three coefficients for the selection variables were statistically significant at the p<0.001 level, together accounting for 14% of the variance in the CSA scores (as indicated by both the adjusted and unadjusted R^2 values). As the standardised coefficients in Table 3 imply, all three selection measures make roughly equal independent contributions to the prediction of CSA score.

INSERT TABLE 3 AROUND HERE

Finally, the path model presented in Figure 1 was estimated. Initially a unidimensional confirmatory factor analytic (CFA) model was tested, with English language fluency as a latent variable with the four IELTS subtest scores as ordinal indicators. The model showed acceptable to good fit according to the CFI value of 0.93, although the TLI was somewhat lower at 0.81 and the RMSEA 0.11 slightly higher than conventional guidelines for goodness

of model fits.(31,37) However, previous methodological research suggests where a latent variable model has categorical indicators CFI may serve as a more accurate estimate of fit than TLI or RMSEA. (38) Once a model for English fluency was developed the remaining SEM was built and tested. The estimated coefficients can be seen in Figure 3 and also Table 4. The model showed a generally acceptable fit to the data with a CFI of 0.93, a slightly lower TLI of 0.88 and a RMSEA of 0.11. As can be seen from Figure 3 SJT scores were more strongly related, in this sample of doctors, to CSA than CPST performance. However, both CPST and SJT scores appeared fairly strongly related to language fluency. Therefore the specific indirect effects were also estimated in order to explore in more detail the putative mediating pathways between language fluency and CSA performance. The relevant coefficients for the indirect pathways are also shown in Table 4. As can be seen in Table 4, the indirect relationship between language fluency and CSA score appears to be mainly mediated by performance on the SJT.

INSERT FIGURE 3 AROUND HERE

DISCUSSION

Key findings and comparison with previous studies

In this study we were able to demonstrate a relationship between the two main selection measures (SJT and CPST scores) and the outcome (CSA performance). In this sample of international medical graduates we observed a relatively stronger relationship between the SJT scores and CSA performance than between CPST and the outcome. Moreover, we were able to provide a detailed analysis of how the relationship between language fluency and CSA rating is mediated by the two differing selection measures, with SJT performance appearing to be the main conduit.

To our knowledge this is the first study to explore how performance at an applied knowledge test and SJT are related to performance at a high fidelity clinical simulation in international medical graduates. However, our findings are broadly in keeping with those previously reported in the literature on medical postgraduate differential attainment. For example, the prior report that high discrepancies in pass rates for the CSA persisted despite controlling for the effects of performance on the AKT suggest that other factors, beyond clinical knowledge, lay behind these.(4) The detailed linguistic study of candidates taking the CSA also highlighted subtle cultural and communication factors as potentially explaining much of the differential pass rate between home and international graduates.(12) Likewise the importance of language fluency in postgraduate medical educational performance has

previously been highlighted in relation differential attainment.(3,5,12) Performance at IELTS was also reported to be a predictor of Fitness to Practice (FtP) events in international medical graduates registered to practice in the UK, though the relationship was more complex than simply better subtest scores reducing the risks. (39) It was also interesting to note that, although females tend to outperform males on SJTs(40), in this sample of overseas doctors we did not observe this sex difference. This may have been because the advantage of being female was not apparent where there were also linguistic and cultural challenges to candidates sitting the SJT.

Strengths and limitations of the study

This was a large and relatively complete dataset on a cohort of international medical graduates applying for a national GP training scheme. By linking data from a variety of sources we were able to explore the relationship between a number of important constructs. Our use of SEM also allowed us to delineate the relationship between language fluency, the two selection measures and the outcome of interest in more detail than a conventional multivariable linear regression. Nevertheless there are several limitations of this study that should be noted. Firstly, CSA performance could not be observed in those who were not selected into GP training, or who had not yet taken the exam. This was have led to some degree of attenuation in the degree of correlation observed between the predictors and the outcome of interest. (41) However, it is reasonable to assume that the restriction of range would have applied to both the main predictors of interest (SJT and CPST scores) and therefore there relative contributions of each to CSA performance are likely to be realistic. There were also challenges in deciding how the CSA result was to be defined. Since trainees are allowed a maximum of four sittings of the CSA the first attempt may not be a true reflection of typical performance, which we used in this study. However, other approaches such as using the average of examination grade are equally, if not more, problematic. Subsequent scores are usually higher, yet it has been shown that, with increasing postgraduate medical exam attempts, chance plays a greater role.(42) These points suggest it is more meaningful to use first attempt data, as other researchers have done previously.(22) Future research could explore other potential predictors, such as number of attempts required to pass: the number of attempts at the PLAB exam has been previously reported to predict the risk of later malpractice in international medical graduates.(39)

A separate potential limitation relates to the measures themselves. Between 2008 and 2012 there have been changes to the GP selection process, as well as with CSA assessments. The weighting of the CPST and SJT in the final selection score has increased since 2008.

Moreover, IELTS requirements changed over the period from a requirement of a minimum score of 7.0 overall to a minimum score of 7.0 on all components. Subsequently this value has been raised again to a minimum overall score of 7.5 (or equivalent) for international medical graduates to be eligible to take the PLAB exam.(43) Thus, these current findings may not fully generalize to subsequent cohorts. Finally, at the time of the study we did not have access to information about the training interventions and programs on which GP trainees were placed. For example, we could not ascertain whether borderline candidates systematically received poorer quality training than their counterparts, which theoretically could account for some effects reported here. For example, weaker trainees could have been assigned to less desirable training posts where learning potential was restricted. Thus, the present dataset did not allow an analysis of the impact of training interventions or differences in training programs, which could also influence outcomes. Further analysis of training interventions in each program would help to understand the 'value-added' of various education and training interventions.

Interpretation of findings

It is relatively clear that the IELTS evaluates language (though other aspects of cognitive ability are no doubt tested, given the academic nature and purpose of the test) and the CPST assesses applied clinical knowledge. In this case it was the IELTS *reading* and *listening* subtest scores that appeared most closely associated with subsequent CSA performance, suggesting that language comprehension is a key ability in relation to actual clinical performance. In contrast to the IELTS and CPST, the constructs evaluated by SJTs are less clear and will vary across contexts. One way of conceptualising the type of SJT used in GP selection is to consider it a special case of a knowledge test. That is, the SJT is measuring the ability to identify and rate professional behaviours within a specific cultural context (i.e. that of a UK health service). It may also be that SJTs can, to some extent in certain contexts, measure '*implicit trait policies*' (ITPs). These can be conceptualised as the beliefs that an individual holds about the attitudes and behaviours they consider important to effective workplace practice. For example, it has previous been reported that individuals who have high level of 'agreeableness' as a trait often consider this to be an important characteristic to exhibit in the workplace. Such ITPs are influenced by work and general personal experience.(44,45) In this way, SJTs can be hypothesised to be measuring the knowledge of interpersonal competence and an understanding of what are effective workplace behaviours. Given that the CSA is designed to measure domains such as data gathering, clinical management and interpersonal skills, it is reasonable to assume that there should be overlap in terms of the constructs measured by the SJT and CSA. Such knowledge of interpersonal competence could be thus considered a necessary, but not

sufficient, condition for subsequent effective performance in the simulated patient-doctor interactions rated in the CSA. According to our results, such knowledge of interpersonal competency appears somewhat distinct from the medical knowledge captured by the CPST.

Our findings also suggest that the relationship between language fluency and other types of knowledge, in this context, is relatively complex. The results highlight how an otherwise competent doctor (who has a good understanding of clinical issues) may have issues with "translating" this semantic medical knowledge into practice. High demands are placed on medical physicians practicing in their own language, and those practising in a second language have the added burden of understanding, especially colloquialisms. (46) It is also difficult to disentangle language from culture. Thus both language fluency and interpersonal competence are needed to make effective behavioural responses when engaging in culturally-laden social interactions. It is these constructs which are likely to account for at least a substantial portion of the 'dark variance' of differential attainment not explained by academic performance. (10) We propose that these findings are not unique to the UK, whereby international medical graduates applying for licensure in other countries are likely to face the same challenges.

In terms of policy and practice our findings imply that if deficits relating to interpersonal competence are identified during selection they are likely to persist, manifesting as poorer performance in the CSA component of the MRCGP exam. Therefore, the selection methods could be useful tools for identifying trainees at risk of poor subsequent performance so that remediation can be offered. Our results also imply that support for international medical graduates should focus on building cultural and interpersonal competence and confidence. An increased focus on evaluating such abilities should also be placed on the tests required of doctors who wish to register to practice in a particular country. Indeed, in the UK the addition of an SJT evaluating knowledge of professionalism within a UK health service context is being considered.(47) Interestingly, the lack of an observed sex difference in SJT performance in the study sample suggests that there would probably be minimal impact on the gender balance of selected international doctors were this approach introduced.

Conclusions

Our findings build on the past research exploring issues relating to the issue of language in the MRCGP assessments.(3,4) We go beyond previous work by providing a more fine-grained analysis where we treat the CSA as a continuous outcome variable in a modelling approach designed to tease out more subtle relationships between the putative predictors. Consequently, we have added to the understanding of both the possible theoretical basis of

differential attainment and where the key causal issues may lie. Thus, we would recommend that efforts directed at reducing these group differences should focus on 'socio-linguistic' factors, rather than clinical knowledge.

References

1. Royal College of General Practitioners. The MRCGP Exam 2017 [Internet]. 2017 [cited 2017 Jan 31]. Available from: http://www.rcgp.org.uk/training-exams/mrcgp-exams-overview.aspx.
2. Wakeford R. International medical graduates' relative under-performance in the MRCGP AKT and CSA examinations. Educ Prim Care2. 2012;23(3):148–52.
3. McManus I, Wakeford R. PLAB and UK graduates' performance on MRCP (UK) and MRCGP examinations: data linkage study. BMJ. 2014;348:g2621.
4. Esmail A, Roberts C. Academic performance of ethnic minority candidates and discrimination in the MRCGP examinations between 2010 and 2012: analysis of data. BMJ. 2013;347:f5662.
5. The High Court of Justice. Neutral Citation Number: [2014] EWHC1416 (admin). 2014.
6. Woolf K, Mcmanus IC, Potts HWW, Dacre J. The mediators of minority ethnic underperformance in final medical school examinations. Br J Educ Psychol. 2013;83(1):135–59.
7. Wakeford R, Denney M, Ludka-Stempien K, Dacre J, McManus IC. Cross-comparison of MRCGP & MRCP(UK) in a database linkage study of 2,284 candidates taking both examinations: assessment of validity and differential performance by ethnicity. BMC Med Educ. 2015;15(1):1–12.
8. 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(apr16_4):g2622.
9. Patterson F, Denney M-L, Wakeford R, Good D. Fair and equal assessment in postgraduate training? A future research agenda. Br J Gen Pract. 2011;61(593):712–3.
10. McManus IC, Dewberry C, Nicholson S, Dowell JS, Woolf K, Potts HWW. Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student selection: meta-regression of six UK longitudinal studies. BMC Med. 2013;11:243.
11. Patterson F, Knight A, Stewart F, MacLeod S. How best to assist struggling trainees?

Developing an evidence-based framework to guide support interventions. Educ Prim Care. 2013 Sep;24(5):330–9.
12. 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. 2014.
13. Personal Social Services Research Unit. Unit Costs of Health & Social Care 2014. 2014.
14. Rimmer A. One in eight GP training posts vacant, despite unprecedented third round of recruitment. BMJ. 2014;349:g6478.
15. Davison I, McManus C, Taylor C. Evaluation of GP Specialty Selection. 2016.
16. Patterson F, Ferguson E, Lane P, Farrell K, Martlew J, Wells A. A competency model for general practice: implications for selection, training, and development. Br J Gen Pract. Royal College of General Practicioners; 2000;50(452):188–93.
17. Patterson F, Tavabie A, Denney M, Kerrin M, Ashworth V, Koczwara A, et al. A new competency model for general practice: implications for selection, training, and careers. Br J Gen Pract. British Journal of General Practice; 2013 May 1;63(610):e331-8.
18. 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(1):50–7.
19. Lievens F, Patterson F. The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. J Appl Psychol. 2011;96(5):927–40.
20. GMC. Knowledge of English - PLAB candidates [Internet]. Available from: https://www.gmc-uk.org/doctors/registration_applications/27047.asp
21. Patterson F, Ashworth V, Zibarras L, Coan P, Kerrin M, O'Neill P. Evaluations of situational judgement tests to assess non-academic attributes in selection. Med Educ. 2012 Sep;46(9):850–68.
22. Patterson F, Lievens F, Kerrin M, Munro N, Irish B. 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):734–41.
23. Chan D, Schmitt N. Situational Judgment and Job Performance. Hum Perform. 2002;15(February 2014):233–54.
24. Kanfer R, Ackerman PL. Motivation and cognitive abilities: an integrative/aptitude-treatment interaction approach to skill acquisition. J Appl Psychol. American Psychological Association; 1989;74(4):657–90.
25. General Medical Council. The state of medical education and practice in the UK.

2016.
26. Patterson F, Kerrin M, Baron H, Lopes S. Exploring the Relationship between General Practice Selection Scores and MRCGP Examination Performance. 2015.
27. IELTS. How Your Score Has Been Calculated? [Internet]. 2018. Available from: http://idpielts.me/results/how-your-score-has-been-calculated/
28. Brown. Confirmatory Factor Analysis for Applied Research. London: The Guilford Press; 2006. 475 p.
29. Hu L, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Model a. 1999;
30. Tuker LR, Lewis C. The reliability coefficient for maximum likelihood factor analysis. Psychometrika. 1973;38:1–10.
31. Browne MW, Cudeck R. Sociological Methods &. Sociol Methods Res. SAGE PERIODICALS PRESS; 1992 Nov 1;21(2):230–58.
32. Kline P. The handbook of psychological testing. London: Routledge; 2000.
33. MacKinnon DP, Lockwood CM, Williams J. Confidence Limits for the Indirect Effect: Distribution of the Product and Resampling Methods. Multivariate Behav Res. 2004 Jan;39(1):99–128.
34. StataCorp. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP; 2015.
35. Muthén LK, Muthén BO. Mplus Version 8. Los Angeles: CA: Muthen & Muthen; 2017.
36. Tabachnik, B. G., Fidell LS. Using multivariate statistics. 4th Editio. Needham Heights, MA: Allyn & Bacon.; 2001.
37. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum Associates; 1988.
38. Yu C-Y. Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes. PhD Thesis. 2002;
39. Tiffin P, Paton LW, Mwandigha LM, McLachlan JC, Illing J. Predicting fitness to practise events in international medical graduates who registered as UK doctors via the Professional and Linguistic Assessments Board (PLAB) system: a national cohort study. BMC Med. 2017;15:66.
40. Whetzel D, McDaniel M, Nguyen N. Subgroup differences in situational judgment test performance: A meta-analysis. Hum Perform. 2008;21:291–309.
41. McManus IC, Dewberry C, Nicholson S, Dowell JS. The UKCAT-12 study: educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a cross-sectional collaborative study of 12 UK medical schools. BMC Med. BMC Medicine; 2013;11:1–25.
42. McManus I, Ludka K. Resitting a high-stakes postgraduate medical examination on

multiple occasions: nonlinear multilevel modelling of performance in the MRCP(UK) examinations. BMC Med. BioMed Central; 2012 Dec 14;10(1):60.

43. GMC. Our required IELTS scores are changing on 18 June 2014. 2014.
44. Motowidlo SJ, Beier ME. Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a situational judgment test. J Appl Psychol. American Psychological Association; 2010;95(2):321–33.
45. Motowidlo S. SJ, Hooper AC, Jackson HL. Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. J Appl Psychol. 2006;91(4):749–61.
46. Patterson F, Cousans F, Coyne I, Jones J, Macleod S, Zibarras L. A preliminary investigation to explore the cognitive resources of physicians experiencing difficulty in training. BMC Med Educ.
47. GMC. Review of the GMC's PLAB test: final report. Manchester; 2014.

Table 1. Sample descriptive statistics and correlations between age, IELTS scores, Applied Knowledge Test (AKT), selection assessment scores (CPST- Clinical Problem Solving Test and SJT- Situational Judgment Test) and performance on the Clinical Skills Assessment (CSA). Note: all correlations are significant at the p<0.01 level except where indicated by an asterisk (*).

		N	Mean	SD	1	2	3	4	5	6	7	8
1.	IELTS reading	3750										
2.	IELTS speaking	3750			0.24							
3.	IELTS writing	3750			0.19	0.25						
4.	IELTS listening	3748			0.40	0.23	0.17					
5.	AKT	2109	10.83	8.50	0.17	-0.01*	0.04*	0.06*				
6.	CPST	3709	237.37	33.03	0.11	-0.04*	0.04*	0.03*	0.42			
7.	SJT	3709	233.98	27.25	0.29	0.13	0.15	0.18	0.18	0.23		
8.	Age	3750	34.09	4.62	-0.13	0.01*	0.01*	-0.16	-0.10	-0.08	-0.14	
9.	CSA	1874	-4.82	10.53	0.12	0.11	0.06*	0.09	0.19	0.13	0.10	-0.15

Predictor	Coefficient	Lower 95% confidence limit	Upper 95% confidence limit	p
Male sex	-5.96	-6.88	-5.03	<0.01
Age	-0.74	-0.84	-0.64	<0.01
SJT score at pass	0.11	0.09	0.13	<0.01
CPST score at pass	0.07	0.06	0.09	<0.01
IELTS reading	1.55	1.24	1.86	<0.01
IELTS speak	1.16	0.80	1.52	<0.01
IELTS writing	0.46	0.12	0.80	0.01
IELTS listening	1.66	1.35	1.98	<0.01
IELTS overall score	5.90	4.88	6.93	<0.01
AKT score at pass	0.36	0.30	0.41	<0.01

Table 2. Results from univariable regressions predicting Clinical Skills Assessment score (at first attempt) from the selection measures, Applied Knowledge Test performance, age and sex.

Predictor	Coefficient (standardised)	LL	UL	p
SJT score at pass	0.07 (0.18)	0.05	0.09	<0.001
CPST score at pass	0.05 (0.17)	0.04	0.07	<0.001
IELTS overall score	4.21 (0.18)	3.18	5.25	<0.001

Table 3. The results of a multivariable linear regression examining the prediction of the components of the IELTS and the two selection measures to independently predict subsequent performance (at first attempt) in the Clinical Skills Assessment (N=1874).

Path	Coefficient	SE	Standardised β
Direct effects			
English fluency → CPST	24.67	1.37	0.44
English fluency → SJT	29.287	1.40	0.63
CPST → CSA	0.05	0.01	0.17
SJT → CSA	0.10	0.01	0.26
Indirect specific effects			
English fluency → CPST → CSA	1.32	0.20	0.07
English fluency → SJT → CSA	2.96	0.30	0.17
Total overall indirect effects			
English fluency→ CSA	2.96	0.30	0.24

Table 4. Results from a path analysis evaluating the direct and indirect relationships between the variables in the structural equation model (see Figure 3). All paths were statistically significant at the p<0.001 level. CPST = Clinical Problem Solving Test; SJT = Situational Judgement Test; CSA = Clinical Skills Assessment. 'English fluency' is conceptualised as a latent variable with the four IELTS subtest scores (reading, speaking, writing and listening) as indicators.

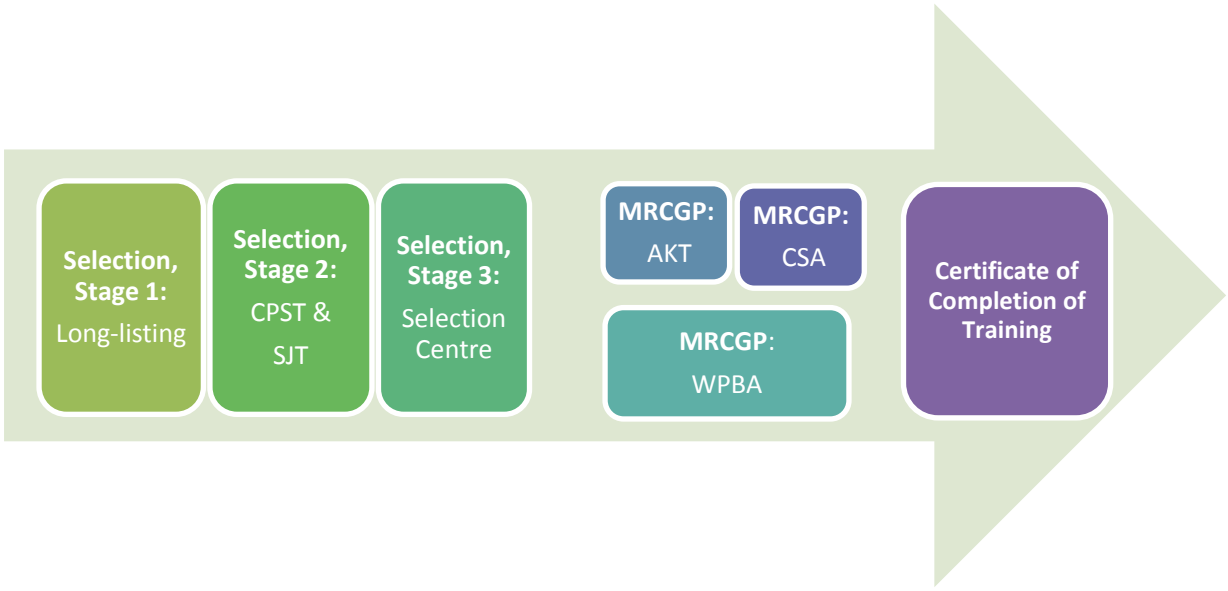

Figure 1. General Practice Selection and MRCGP Assessment Pathway. MRCGP = Membership of the Royal College of General Practitioners; CPST = Clinical Problem Solving Test; SJT = Situational Judgement Test; WPBA = Workplace Based Assessment; AKT = Applied Knowledge Test; CSA = Clinical Skills Assessment.

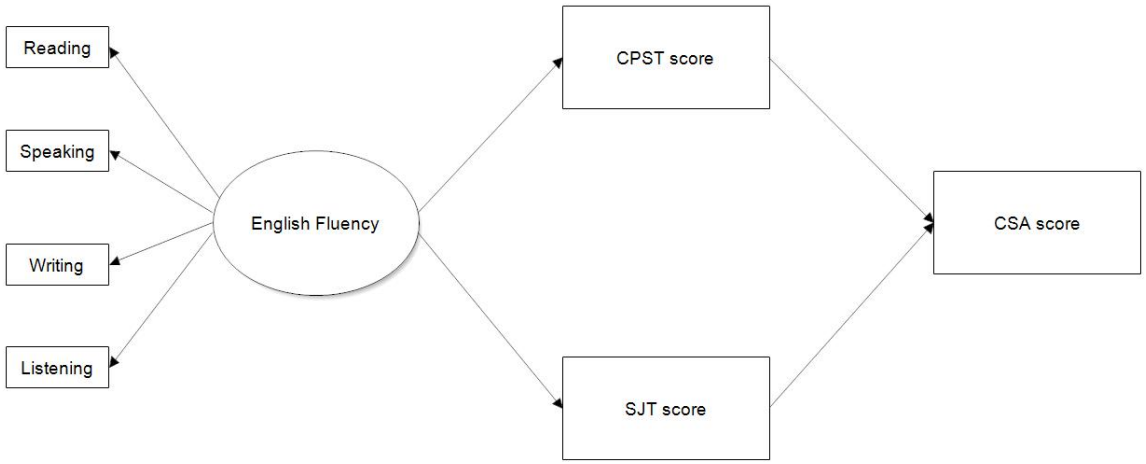

Figure 2. Proposed model examining the relation between language fluency, performance on the selection assessments and subsequent Clinical Skills Assessment (CSA) performance.

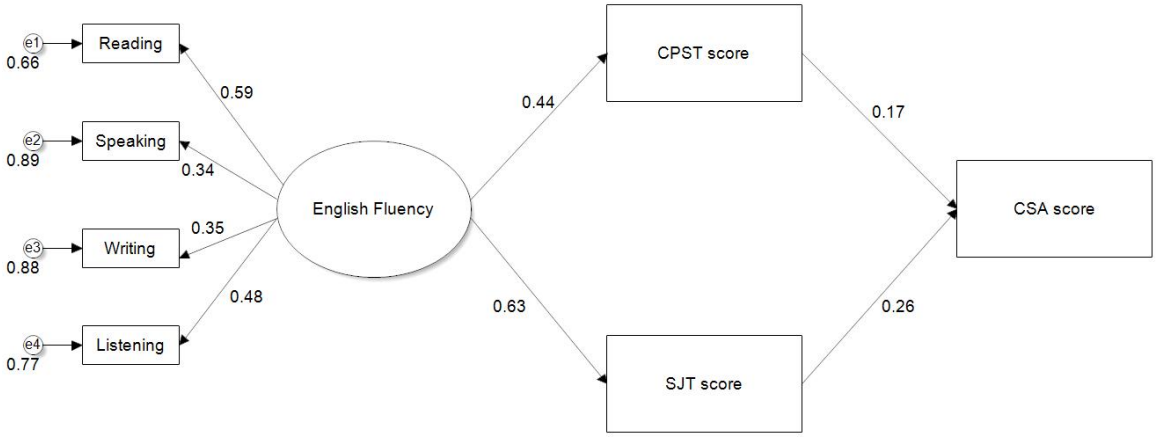

Figure 3. Final model examining effects of language ability and knowledge on CSA performance (N=1874). Note: 'English fluency is conceptualised as a latent variable with the four IELTS subtest scores (reading, speaking, writing and listening) as indicators.