

This is a repository copy of *Situational judgement test performance and subsequent misconduct in medical students*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/184825/>

Version: Published Version

Article:

Tiffin, Paul Alexander orcid.org/0000-0003-1770-5034, Sanger, Emily, Smith, Daniel et al. (2 more authors) (2022) Situational judgement test performance and subsequent misconduct in medical students. *Medical Education*. ISSN 0308-0110

<https://doi.org/10.1111/medu.14801>

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

RESEARCH ARTICLE

Situational judgement test performance and subsequent misconduct in medical students

Paul A. Tiffin^{1,2}  | Emily Sanger³  | Daniel T. Smith⁴  | Adam Troughton⁴  | Lewis W. Paton^{1,2} 

¹Department of Health Sciences, University of York, York, North Yorkshire, UK

²Hull York Medical School, York, North Yorkshire, UK

³Leeds and York Partnership NHS Foundation Trust, Leeds, Yorkshire, UK

⁴Education and Standards, General Medical Council, London, UK

Correspondence

Paul A. Tiffin, Department of Health Sciences, University of York, York, North Yorkshire, UK.
Email: paul.tiffin@york.ac.uk

Funding information

The UCAT Board, Grant/Award Number: Not applicable; UCAT Board; NIHR, Grant/Award Number: CDF-2015-08-011

Abstract

Introduction: Situational judgement tests (SJTs) have been widely adopted, internationally, into medical selection. It was hoped that such assessments could identify candidates likely to exhibit future professional behaviours. Understanding how performance on such tests may predict the risk of disciplinary action during medical school would provide evidence for the validity of such SJTs within student selection. It would also inform the implementation of such tests within student recruitment.

Methods: This cohort study used data for 6910 medical students from 36 UK medical schools who sat the University Clinical Aptitude Test (UCAT) SJT in 2013. The relationship between SJT scores at application and the risk of subsequent disciplinary action during their studies was modelled. The incremental ability of the SJT scores to predict the risk of disciplinary action, above that already provided by UCAT cognitive test scores and secondary (high) school achievement, was also evaluated in 5535 of the students with information available on this latter metric.

Results: Two hundred and ten (3.05%) of the students in the cohort experienced disciplinary action. The risk of disciplinary action reduced with increasing performance on the admissions SJT (odds ratio (OR) 0.80, 95% confidence interval (CI) 0.69 to 0.92, $p = 0.002$). This effect remained similar after adjusting for cognitive performance and prior academic attainment (OR 0.77, 95% CI 0.65 to 0.92, $p = 0.004$). The overall estimated effect-size was small (Cohen's $d = 0.08$) and no evidence of 'threshold' effects were observed for the SJT scores and risk of disciplinary action.

Conclusions: Performance on admissions SJTs can, at least modestly, incrementally predict the risk of subsequent disciplinary action, supporting their use in this context. However, for this SJT and outcome, there did not seem a distinct threshold score above which the risk of disciplinary action disproportionately increased. This should be considered when using the scores within medical selection.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Medical Education* published by Association for the Study of Medical Education and John Wiley & Sons Ltd.

1 | INTRODUCTION

Medical regulators expect physicians to behave professionally¹ with serious lapses in conduct adversely impacting patients, the public and the profession. Such expectations extend to medical students,² where professionalism lapses during undergraduate training are associated with subsequent conduct issues in medical practise.^{3,4} Moreover, addressing professionalism issues following medical school entry is challenging.⁵ Therefore, selection processes should evaluate personal, 'non-academic', qualities relevant to ethical practise.⁶ In this context 'non-academic abilities' refers to qualities relevant to interpersonal functioning not directly related to traditional concepts of intellectual ability.

However, measuring such personal attributes in high-stakes situations is challenging, where 'faking' and coaching effects may influence performance.⁷ Nevertheless, situational judgement tests (SJTs) are considered a cost-effective and valid approach to assessing non-academic abilities in medical selection.⁸ In SJTs applicants are presented with a scenario and evaluate possible responses to the situation, commonly ranking or rating them in relation to perceived appropriateness. SJTs have been rapidly implemented internationally, though a meta-analysis highlighted a relative paucity of evidence for their use in undergraduate, versus postgraduate, medical selection. Moreover, on average, the validity coefficients reported for SJT scores used to select into medical school were substantially lower than those deployed in postgraduate recruitment.⁹ Whilst the test content varies, most SJTs used for medical school selection include items evaluating candidates' understanding of professional, ethical or moral practice. For example, the Medical College Admission Test (MCAT) SJT aims to predict 'student professionalism'.¹⁰ Thus, it could be hypothesised that better performing candidates on such selection assessments are at lower risk of subsequent misconduct.

The UCAT was introduced into medical selection in 2006 and is used by most UK-based medical schools and, more recently, a number of Australasian institutions.¹¹ The SJT component of the UCAT, introduced in 2013, was text based with a rating style response format. Content covered the domains of 'integrity', 'perspective taking' and 'team working'. An earlier validation study reported modest, though statistically significant, correlations between the scores and tutor ratings for those subsequently entering medical school.¹² Most of the UCAT test-takers in 2013 who entered medical school have now completed their undergraduate studies. These test data, as well as any disciplinary events that occurred during medical school, are stored in the UK Medical Education Database (UKMED).¹³ Moreover, only two of the 37 UK medical schools operating in 2013 used the SJT scores (albeit weakly) within the selection process at that time.¹⁴ This would have reduced the direct 'range-restriction' effects for observed SJT scores due to candidate selection. Thus, this presented a valuable opportunity to evaluate to what extent earlier findings related to validity were supported by evidence of the risk of disciplinary action in relation to SJT performance. Moreover, this would be the first study to link performance on an SJT taken prior to medical school entry to subsequent disciplinary events. Previous studies have, to

date, only evaluated this relationship with SJT scores taken after matriculation.¹⁵⁻¹⁷

Thus, the aims of this study were to establish whether performance on the UCAT SJT at application was associated with the subsequent risk of disciplinary action during medical school. A secondary aim was to establish to what extent, if any, the SJT scores demonstrated incremental validity over the other two main selection assessment metrics (cognitive performance and academic achievement)? The findings would have implications for how such SJTs should be optimally implemented within the selection process in order to reduce the risk of professionalism lapses, thus improving patient safety.

2 | METHODS

2.1 | Ethics

As the study used routine, deidentified data ethical approval was not required and this was confirmed in writing by the Chair of York Health Sciences Ethics Committee.

The study data were held in a 'safe-haven' where individual data cannot be extracted.^{18,19} Moreover, any results from the UKMED must be presented in blunted form, with frequencies rounded to the nearest five.^{18,20}

2.2 | Data availability

Data were available for all 15 245 applicants sitting the UCAT in 2013. The flow of data is depicted in Figure 1.

Inclusion criteria:

- The UCAT (including the SJT) was sat in 2013.
- A UK medical school had subsequently been entered.
- At least 4 years of medical school had been completed.

The latter criteria was included to ensure that the 'exposure time' at medical school was similar in all cases. At the time, the data were not available to confirm a fifth year of medical school had been completed. However, a small number of cases (<5) were retained where the student had left medical school before 4 years were completed but had been reported as experiencing a disciplinary event during their studies.

Exclusion criteria:

- One medical school (Glasgow) had not returned annual data relating to fitness to practise referrals at the time of the study. Thus, students entering this institution were excluded.
- One thousand three hundred seventy observations (19.86% of the sample) from doctors who did not have an English or Scottish Advanced school qualification reported were excluded from the multivariable analyses involving adjustment for prior educational attainment.

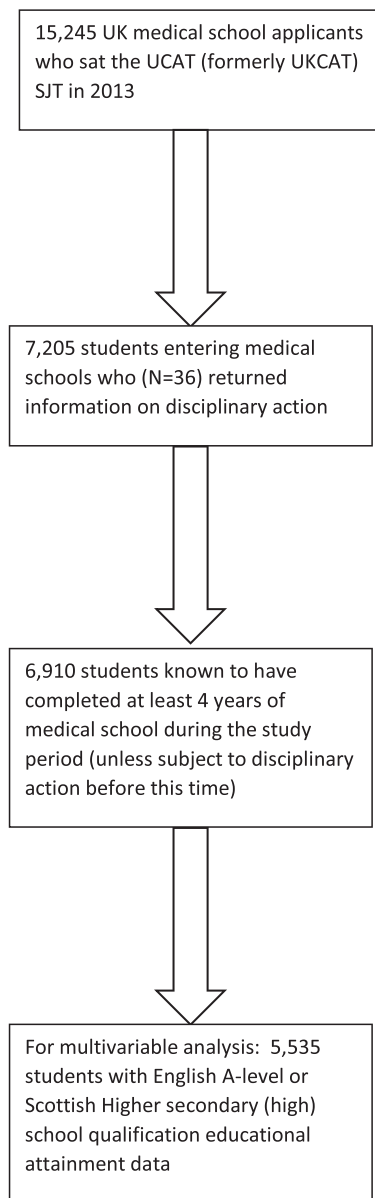


FIGURE 1 The flow of data through the study

2.3 | Data management

2.3.1 | Outcome variable—Referral in relation to fitness to practise

The outcome of interest was the referral of a medical student to ‘fitness to practise’ (disciplinary) processes, as reported by UK medical schools to the General Medical Council (GMC) as the medical regulator. Such referrals are reported annually and data were available from all but one of the 37 UK medical schools.

Fitness to practise referrals reported as solely ‘health’ related were excluded as our focus was on predicting conduct issues. Referrals resulting in ‘no action’ ($n = 25$) were also excluded. Consequently, fitness to practise referrals resulting in some formal action are henceforth described as ‘disciplinary events’. A minority of

students had more than one referral ($n = 65$). As it was not always clear whether the complaints raised were independent of each other, we defined the outcome dichotomously (one or more disciplinary events reported versus none). In the analytic dataset there were 165 students with at least one disciplinary event reported.

In order to explore the disciplinary event reporting process we also compared the agreement between medical school and student-reported disciplinary events. The latter are self-declared by students applying to the GMC for provisional registration. For this evaluation we used all UKMED data. For those graduating in 2019 and 2020 48% of disciplinary events reported by the school ($n = 575$) were also declared by the applicant. Conversely, 69% of events declared by an applicant ($n = 400$) were also declared by the school. The two modes of reporting demonstrated moderate agreement ($\kappa = 0.55$). School-based reporting tended to contain a higher proportion of conduct issues. Conversely, self-declarations were more likely to report lower level conduct and health issues that may not have been subject to official disciplinary processes. Agreement varied substantially across medical schools (κ values 0.08 to 0.95).

2.3.2 | Predictor variables—Demographic variables

As in previous similar research^{21,22} self-reported ethnicity was dichotomised into ‘White’ and ‘non-White.’ Secondary school type attended was dichotomised into state-funded schools and private-funded schools. Socio-economic status was reported, mainly based on the parent or guardian’s occupation, and dichotomised into ‘professional’ versus ‘non-professional’ background. Age was categorised into ‘mature’ (20 or older at UCAT sitting) or ‘non-mature’.

UCAT SJT performance

The equated 2013 UCAT SJT scores were available²³ and transformed into standardised, z-scores (mean 0, SD 1). In the UK UCAT SJT scores are collapsed and reported as four bands to selectors. These bands were calculated according to the method described by Work Psychology Group, the test designers.²⁴

UCAT cognitive test performance

The summed score across all four cognitive scales (‘total UCAT score’) was transformed into a standardised z-score based on the scores for all candidates sitting the test in 2013.²³ Note that this would have included those applying to dental and medical schools.

2.4 | Prior educational attainment

The subject and grade attained at Advanced level (A-level) examinations for those students from England and Wales (who were not graduate entrants) were available. This information was also available for ‘Scottish Higher’ qualifications. A metric for advanced qualifications was derived, using a method previously described, in order to create a z-score for ‘prior educational attainment’.^{21,25}

2.5 | Statistical analysis

The timing of the complaints and disciplinary processes were not recorded consistently. Therefore, 'time to event' modelling (i.e., survival analysis) was not feasible. However, it was possible to achieve a cohort with approximately similar 'exposure times'. This was done by excluding the small number of students who had left medical school without having completed at least 4 years of study (except for those experiencing disciplinary events).

To account for variations (e.g., in disciplinary processes) between medical schools we used generalised estimation equations (GEEs) to link the predictor variables, via a logit link function, to the odds of observing a disciplinary event.

Initially, univariable analyses were performed to understand the raw (unadjusted) relationship between the predictors and the risk of subsequently experiencing a disciplinary event. Multivariable models were then built to evaluate the adjusted relationship between SJT and disciplinary events, controlling for both standardised total UCAT score and prior educational attainment, as well as any variable interactions. Demographic variables were not included in our models because such characteristics are not generally used in UK medical selection. However, the demographic data were used to describe the study cohort and to inform imputation of the simulated outcomes for non-entrants where relevant.

Missing predictor values were not prevalent, and the data were analysed using listwise deletion. Data management and analyses were conducted using STATA MP version 15.²⁶

2.6 | Simulating disciplinary events in non-entrants

In order to, at least crudely, adjust for restriction of range, disciplinary events were simulated via a single imputation. This approach has been previously used to adjust for selection effects when dealing with categorical outcomes.^{11,27} The imputation was performed using chained equations and was informed by the UCAT cognitive scores and available demographic variables.

3 | RESULTS

The flow of data through the study is shown in Figure 1.

Of the 6910 students in the study, 210 (3%) had at least one disciplinary event resulting in some action. The sample sociodemographic characteristics are displayed in Table 1.

3.1 | Predicting disciplinary events

Table 2 displays the model results. These results show that individuals with higher standardised scores on the SJT had statistically significantly lower odds of experiencing a disciplinary event (OR 0.80, 95% CI 0.69 to 0.92, $p = 0.002$). Specifically, on average, for every standard deviation above the mean scored on the SJT, entrants had 20% lower odds of experiencing a disciplinary event. Odds ratios are challenging to interpret where the outcome event is relatively uncommon. Therefore, an approximate effect-size, in terms of Cohen's d , was calculated as 0.08 (i.e., a small effect-size) for an OR of 0.80.²⁸

Those with higher prior educational attainment and UCAT cognitive scores were at lower risk of a disciplinary event being reported (OR 0.70, 0.58 to 0.84, $p < 0.001$). Controlling only for the influence of the advanced qualifications increased the independent predictive effect of the SJT scores. In contrast, controlling for the total UCAT cognitive scale scores modestly weakened this observed relationship. The estimated independent effect of the SJT scores on the odds of a disciplinary event also increased slightly when controlling for both prior educational attainment and the UCAT cognitive scale scores (Table 2). No statistically significant interaction terms were observed. In order to assess for any potential 'threshold' scores, associated with a non-linear increase in disciplinary risk, we evaluated quadratic and cubic terms for the SJT scores. These were not statistically significant. In terms of the UCAT cognitive scales, on univariable analysis, only higher *decision analysis* (OR 0.74, 0.63 to 0.87, $p < 0.001$) and *verbal reasoning* scores (OR 0.82, 0.71 to 0.94, $p = 0.006$) were significantly protective of a disciplinary event.

TABLE 1 Demographic and educational information for applicants included in the analytic dataset (N = 6910)

Demographic variable	Proportion (%)	Missing (%)
Male gender	3025/6910 (43.80%)	0/6910 (0%)
'Non-White' ethnicity	2065/6060 (34.06%)	850/6910 (12.29%)
Attended state school	4340/5785 (75.02%)	1125/6910 (16.27%)
UK resident	6190/6910 (89.62%)	0/6910 (0%)
Non-professional background	250/6910 (3.63%)	0/6910 (0%)
Age >20 at UCAT sitting	1180/6910 (17.11%)	0/6910 (0%)
Educational variable	Standardised mean (SD)	Missing (%)
UCAT SJT z-score	0.30 (0.87)	0/6910 (0%)
UCAT total cognitive z-score	0.70 (0.74)	0/6910 (0%)
Advanced qualification z-score (compared to applicants)	0.05 (0.61)	1380/6910 (19.99%)

TABLE 2 Odds ratios from generalised estimating equation (GEE) models predicting if a medical student was subject to at least one disciplinary event, according to their UCAT performance and prior educational achievement ('advanced qualifications': AQs), expressed as a standardised z-scores

Variable	Unadjusted OR (95% CI)	p	OR adjusted for advanced qualifications only (95% CI)	p	OR adjusted for UCAT cognitive total score only (95% CI)	p	OR adjusted for both advanced qualifications and UCAT cognitive score (95% CI)	p
Standardised SJT score	0.80 (0.69 to 0.92)	0.002	0.73 (0.62 to 0.87)	<0.001	0.83 (0.72 to 0.97)	0.017	0.77 (0.65 to 0.92)	0.004
AQ z-scores	0.76 (0.63 to 0.91)	0.003	0.76 (0.63 to 0.92)	0.004	Not applicable	-	0.82 (0.67 to 1.00)	0.048
Standardised UCAT total cognitive score	0.70 (0.58 to 0.84)	<0.001	Not applicable	-	0.73 (0.61 to 0.88)	0.001	0.71 (0.57 to 0.90)	0.004

Approximately 20% of the students ($n = 1380$) did not have English or Scottish high school qualifications reported, of which around half ($n = 720$) were graduate entrants. As a sensitivity analysis we repeated the univariable analyses for only those with advanced school achievement data. Consequently, the univariable relationship between SJT score and the risk of a disciplinary event increased modestly (OR 0.73, 0.62 to 0.87, $p < 0.001$).

The results of the modelling between the SJT score band achieved and the risk of experiencing a disciplinary event are shown in Table 3. In this regard, on univariable analysis, there was a significant decreased odds of a disciplinary event for those scoring in band 1 (highest) compared with the other bands. The relationship between the SJT score bands and the odds of a disciplinary event remained similar in magnitude once adjustment was made for the effects of the other two selection metrics. However, there were slightly fewer observations as only students with advanced qualification data recorded were included. Consequently, only the differences between those in band 4 and band 1 remained statistically significant at the $p < 0.05$ level in this multivariable model. Table 3 also contains a column with the actual (blunted) numbers and proportions of students in each SJT score band with at least one disciplinary action recorded. Some medical applicants achieving only band 3 or 4 SJT scores would have failed to enter medical school due to direct and indirect selection effects. Therefore, in order to more realistically convey the potential impact of using the SJT bands for 'screening out' low scoring medical school applicants a separate column in Table 3 is included. This includes imputed (simulated) disciplinary events for those who did not enter medical school. Accordingly, it is apparent that using band 4 as a screening threshold, in order to reject 145 candidates who would be likely to face at least one disciplinary event, we would need to reject 1380 'low risk' candidates. This ratio of approximately 10 (1380/145) is termed the 'number needed to reject' (NNR).^{11,29} It represents the number of 'low risk' candidates needed to be rejected in order to screen out one 'high risk' applicant. In contrast, the NNR for band 3 is around 17.

The SJT scores can be conceptualised as a 'screening test'. Therefore, we calculated the area under the curve (AUC) for the receiver operator characteristic (ROC) curve predicting the risk of a disciplinary event for various score thresholds (Figure 2). The AUC was 0.58 (95% confidence interval 0.53 to 0.62). Thus, whilst statistically significantly different from 0.5 (i.e., 0.5 representing a test no better than chance) it was a relatively low value. Moreover, the ROC curve lacks an obvious 'elbow' that could represent an optimum cut-point.

4 | DISCUSSION

We observed that higher UCAT SJT scores at application were associated with a reduced odds of at least one disciplinary event occurring during undergraduate medical study. When adjusting for the influence of overall performance on the cognitive components of the UCAT and prior educational attainment this relationship remained statistically significant, though the actual effect-size was modest.

TABLE 3 The risk of disciplinary action depending on the scoring band achieved in the UCAT SJT for both medical school entrants (N = 6910) and all medical school applicants sitting the test in 2013 (N = 15,245)

SJT score band comparison	Proportions (%) with disciplinary action (entrants only)	Proportions (%) with disciplinary action (entrants and also simulated outcomes for non-entrants)	Unadjusted OR (95% CI) ^a	p	OR adjusted for both UCAT score and AQ z-score (95% CI) ^a	p
Band 2 vs. Band 1	95/3075 (3.15%) vs. 40/1975 (1.97%)	240/5950 (4.05%) vs. 90/3275 (2.75%)	1.41 (1.03 to 1.93)	0.03	1.43 (0.97 to 2.11)	0.072
Band 3 vs. Band 1	60/1640 (3.78%) vs. 40/1975 (1.97%)	245/4500 (5.49%) vs. 90/3275 (2.75%)	1.50 (1.06 to 2.23)	0.02	1.48 (0.96 to 2.28)	0.075
Band 4 vs. Band 1	15/215 (5.99%) vs. 40/1975 (1.97%)	145/1525 (9.51%) vs. 90/3275 (2.75%)	2.29 (1.26 to 4.16)	0.006	2.55 (1.28 to 5.08)	0.008
Band 3 vs. Band 2	60/1640 (3.78%) vs. 95/3075 (3.15%)	245/4500 (5.49%) vs. 90/3275 (2.75%)	1.06 (0.80 to 1.42)	0.68	1.03 (0.74 to 1.46)	0.844
Band 4 vs. Band 2	15/215 (5.99%) vs. 95/3075 (3.15%)	145/1525 (9.51%) vs. 90/3275 (2.75%)	2.16 (0.93 to 2.83)	0.09	1.79 (0.95 to 3.35)	0.071
Band 4 vs. Band 3	15/215 (5.99%) vs. 60/1640 (3.78%)	145/1525 (9.51%) vs. 245/4500 (5.49%)	1.52 (0.86 to 2.72)	0.15	1.73 (0.91 to 3.29)	0.097

Note: In the latter case, disciplinary events were simulated using data imputation for those applicants not entering a UK medical school included in the study.

^aCalculated for actual disciplinary events only.

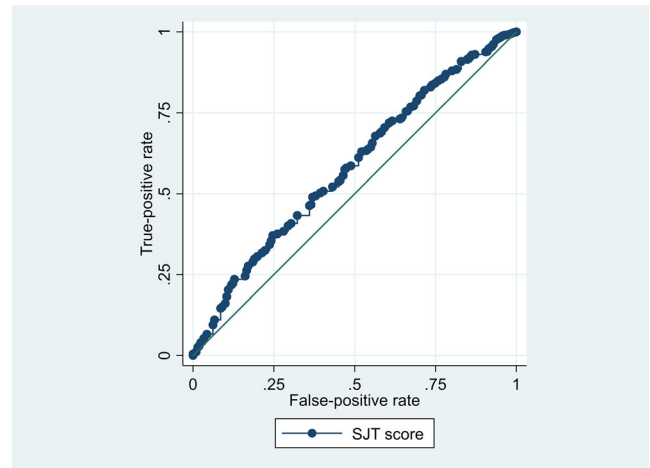


FIGURE 2 The receiver operator characteristic (ROC) curve predicting the risk of a disciplinary event for various thresholds of the UCAT Situational Judgement Test (SJT) score

Our findings add to those from the smaller, original validity study for the UCAT SJT which reported that scores correlated with supervisor ratings across a number of domains, including (perceived) 'integrity'.¹² Our univariable findings are also consistent with those reported by a previous study that found higher SJT scores in second-year medical students were associated with fewer professionalism concerns.¹⁷ However, this latter study used an SJT taken after admission to medical school. Likewise, two previous studies, both using data from the UKMED, explored the relationship between performance on an SJT taken after medical school entry and disciplinary action.^{15,16} These studies reported univariable relationships between the SJT scores and misconduct in both medical students¹⁵ and during the first 5 years of practice.¹⁶ However, in both cases, these relationships were no longer statistically significant after adjusting for the influence of educational and demographic factors. Neither study reported results solely adjusted for performance on other selection metrics. Thus, their multivariable findings are not directly comparable to the present ones. Moreover, the latter study only included 65 doctors with disciplinary action against them. Consequently, the study would have been underpowered to adequately test a multivariable model. Nevertheless, it is noteworthy that the magnitude of the relationship between SJT performance and disciplinary events (0.84, 0.62 to 1.13, $p = 0.24$) was similar to that observed in the present study.

In relation to medical school selection SJTs: a Canadian study reported statistically significant correlations between performance on a video-based SJT and performance on subsequent examinations which included content related to knowledge of medical ethics.³⁰ However, the SJT had been administered in a low-stakes context and incremental validity was not evaluated.

In the context of existing evidence, the unique contributions of the present study has been to show that the scores from SJT-format assessments, administered prior to medical training, may be incrementally predictive of disciplinary action. The effect-sizes observed were

small (i.e., Cohen's d of approximately 0.08). However, this must be placed in context of the challenges of both predicting an uncommon behavioural event in a highly pre-selected (medical student) population, and the wider selection literature. Indeed, even when concerned with future academic performance, applicant cognitive test scores are only modestly predictive of passing future high-fidelity clinical simulations at first attempt. For example, an odds ratio of 1.34 (Cohen's d approximately 0.16) has been recently reported in this respect.³¹

4.1 | Possible interpretations

The UCAT SJT content domains are labelled 'team working', 'perspective taking' and 'integrity'. Thus, the main underlying construct evaluated by this test may be a candidate's procedural knowledge of interpersonal appropriateness.³² This could be assumed to be a prerequisite of demonstrating professional behaviour. Thus, our findings broadly support the construct validity of the SJT. Moreover, given the costs to stakeholders of selection procedures it was important to assess whether an additional test adds any incremental value.

Conditioning the SJT scores on educational attainment modestly increased their independent predictive validity. This finding was not entirely unexpected. The emotional intelligence literature suggests that once a certain amount of academic or intellectual ability is reached then it is non-academic attributes that are most predictive of interpersonal functioning.³³ Thus, prior educational performance could be considered something of a 'collider' ('reverse cofounder') in this context.³⁴ However, this effect was not observed for UCAT cognitive ability. This may be because the SJT was substantially verbally loaded, demanding a reasonably high level of language comprehension. Therefore the constructs under evaluation may have overlapped to a larger extent than between the SJT and educational attainment. Indeed, we observed that higher scores from two of the more verbally loaded cognitive elements of the UCAT were associated with a reduced risk of a disciplinary event. Indeed, the close link between verbal and social ability has long been recognised.³⁵

The UCAT SJT tends to differentiate best between candidates at the lower end of performance.³² However, the expected non-linear relationship between SJT scores and risk of disciplinary action was not observed. This may have been because our sample was restricted to entrants, who generally scored above average on the SJT. The resulting score distribution may have then obscured any non-linear trend.

4.2 | Strengths and limitations

This study used a national dataset and almost all UK medical applicants sit the UCAT.³⁶ Standardisation of the test scores would have adjusted, to some extent, for the restriction of range that affects selection studies.³⁷ Moreover, in 2013 the SJT scores were barely used in selection, minimising the direct impact on restriction of range. However, some 'indirect range restriction' would have occurred, if

only because SJT scores correlate, to some extent, with other selection metrics. Outcome data were missing from one medical school though this is unlikely to have substantially impacted the overall findings.

Misconduct procedures vary across institutions.³⁸ However, the UK medical regulator and Medical Schools Council provides guidance regarding the nature of misconduct in students that should be reported by universities² and this may have improved consistency. Moreover, few disciplinary referrals resulted in 'no further action' ($n = 25$), suggesting most complaints were relatively serious and well evidenced. Additionally, the use of GEEs would have accommodated any dependency in the observations within medical schools.

Though 'time to event' models were not feasible we attempted to ensure that the time spent at medical school would be similar in all cases. Nevertheless, some minor variation in duration of 'exposure time' would have occurred. For example, some students may have retaken a year. A very small number of students may not have completed the final year as we could not confirm this. However, attrition at this stage of training is very low³⁹ so this is unlikely to have impacted on our findings.

Our sensitivity analysis results suggest that our findings may not generalise to the 20% of students who lacked information on English and Scottish school qualifications. Around half these students were graduate entrants and others may have been schooled outside the United Kingdom. For these groups the relationship between SJT performance and disciplinary events may be weakened or absent.

4.3 | Implications for policy and practice

UK medical schools vary in their use of the UCAT SJT scores in selection. Some screen out candidates scoring in the lowest band ('4'), whilst others allocate a certain number of points to feed into a selection algorithm.¹⁴ Our findings lend some support to the former practice. However, despite the high odds of a disciplinary event for those in band 4, few students actually fell into this low scoring category. This reduces the opportunity to exclude meaningful numbers of applicants at relatively high risk of misconduct. Moreover, given the lack of a clear 'threshold', it may be more practical to report more granular performance metrics to selectors. The UCAT SJT scores are not normally distributed so the use of a non-parametric, ranking-based metric, such as deciles might be appropriate. However, ranking would mean scores for those in the upper deciles would be very similar. The reporting of the continuous scores, as practised in Australasia, could also be justified. However, SJTs, in this context, do not discriminate between individuals with the precision of cognitive tests. Thus, continuous scores may provide an unrealistic impression of their accuracy and should be accompanied by clear guidance to institutions about the optimal ways of using them within selection.

Personnel selection can be conceptualised as a 'pareto-optimisation' situation with better and worse trade-offs.^{40,41} The UCAT SJTs appear less sensitive to certain demographic characteristics compared to academic or cognitive measures. Therefore, their use may facilitate

diversity in medicine.⁴² Consequently, medical schools may consider allowing higher SJT performance to compensate for relatively lower achievement in other domains. Such trade-offs could be justified given that knowledge of interpersonal effectiveness may be at least as important as clinical knowledge when predicting clinical simulation performance.⁴³

Only around half of institutionally reported disciplinary events were also reported at student self-declaration. This could partly be explained by some differences in declaration guidelines for students versus institutions. However, some substantial underreporting by students looks likely. Clearer reporting advice to universities and students may be needed. Moreover, routine cross-validation between the declaration modalities should be performed by the regulator.

4.4 | Directions for future research

Our findings need replication in future cohorts—increased coaching and candidates' familiarity with the SJT format and content could influence its psychometric properties. Evidence of a 'footprint' of reduced actual medical misconduct resulting from SJT-based selection should also be sought, controlling for potentially confounding secular trends.

Most SJTs are still essentially 'paper and pencil' tests. Longer term, increased access to immersive formats (e.g., virtual reality) and the advent of 'computational psychometrics' may improve our ability to evaluate relevant behavioural tendencies.⁴⁴ These could give rise to more effective assessments of interpersonal abilities linked to clinical effectiveness and professionalism.

5 | CONCLUSIONS

Performance on SJTs at medical school application may be independently associated with the risk of subsequent disciplinary action. However, to realise their benefits such assessments must be optimally used within selection. Their effective implementation should ultimately improve both patient safety and quality of care.

ACKNOWLEDGEMENTS

The authors would like to thank Rachel Greatrix (Chief Operating Officer, UCAT) and the members of the UK Medical Education Database (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 UK Medical Education Database (UKMED) [project number P115], with the extract generated on 25th November 2020, reissued 14 October 2021. Approved for publication on 20 December 2021. I am grateful to UKMED for the use of these data. However, UKMED bears no responsibility for their analysis or interpretation. The data includes 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 Copyright Higher Education Statistics Agency

Limited. The Higher Education Statistics Agency Limited 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. PAT was funded via a National Institute for Health Research Career Development Fellowship (NIHR CDF-2015-08-011) and the UCAT Board. This paper thus presents independent research part-funded by the NIHR. The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR or the Department of Health and Social Care.

CONFLICT OF INTEREST

The lead author (PAT) has previously collaborated on research with Work Psychology Group (WPG) who are contracted to design the UCAT Situational Judgement Test. LWP is partly funded by the UCAT Consortium. The UCAT consortium partly funded this research but did not play an active role in determining the study design or reporting the results.

AUTHOR CONTRIBUTIONS

PAT led on the conceptualisation, data management and analysis of the data, and the writing of the report. ES contributed to data management and analysis as well as drafting the article. AT led on evaluation of the agreement between the self and medical school declarations of disciplinary events and critically appraised the draft paper. DS contributed to data management and appraising and write up of the final report. LWP contributed to the data analysis and writing and critically appraising of the report.

ETHICS STATEMENT

As the study used routinely collected, deidentified data, ethical approval was not required. This was confirmed in writing by the Chair of the University of York Health Sciences Ethics Committee. Moreover, individual informed consent for the use of the data was not required for this study. This is because all the data used were held within the UKMED. The use of personal data in the UKMED is not reliant on individual consent from data subjects as it is not a necessary condition for processing under the General Data Protection Regulation (GDPR). This allows personal data to be used without consent where it is necessary for statutory functions. In this instance these powers are granted by the Medical Act (1983). All the data used in the study are held in a 'safe-haven' in which the analysis takes place. In the safe haven arrangement only deidentified data are shared with researchers and only summary reports, not individual data, can be extracted. This prevents leakage of potentially identifiable data (The General Medical Council 2020). To access the data, a project proposal must be approved by the UKMED research subgroup. Any published findings using data from the UKMED must be presented in blunted form in accordance with Higher Education Statistics Agency (HESA) statistical disclosure controls (General Medical Council 2020, Higher Education Statistics Agency 2021). This means all frequencies were rounded to the nearest multiple of 5. This data blunting acts as a further precaution against identifying individuals or small groups of doctors in the study sample.

ORCID

Paul A. Tiffin  <https://orcid.org/0000-0003-1770-5034>

Emily Sanger  <https://orcid.org/0000-0003-0855-2208>

Daniel T. Smith  <https://orcid.org/0000-0003-1215-5811>

Adam Troughton  <https://orcid.org/0000-0003-1568-7581>

Lewis W. Paton  <https://orcid.org/0000-0002-3328-5634>

REFERENCES

- General Medical Council. *Good Medical Practice*. GMC; 2019.
- General Medical Council, Medical Schools Council. Professional behaviour and fitness to practise: guidance for medical schools and their students. 2016. Accessed August 4, 2021. https://www.gmc-uk.org/-/media/documents/professional-behaviour-and-fitness-to-practise-20210722_pdf-66085925.pdf?la=en&hash=C642DC46EDE6CAD7815573F19ACE1390F0D69A50
- Papadakis MA, Teherani A, Banach MA, et al. Disciplinary Action by Medical Boards and Prior Behavior in Medical School. *N Engl J Med*. 2005;353(25):2673-2682. doi:10.1056/nejmsa052596
- Santen SA, Petrusa E, Gruppen LD. The relationship between promotions committees' identification of problem medical students and subsequent state medical board actions. *Adv Health Sci Educ*. 2015;20(2):421-430. doi:10.1007/s10459-014-9536-2
- Papadakis MA, Paauw DS, Hafferty FW, Shapiro J, Byyny RL. Perspective: the education community must develop best practices informed by evidence-based research to remediate lapses of professionalism. *Acad Med*. 2012;87(12):1694-1698. doi:10.1097/ACM.0b013e318271bc0b
- Powis D. Selecting medical students: an unresolved challenge. *Med Teach*. 2015;37(3):252-260. doi:10.3109/0142159X.2014.993600
- Duckworth AL, Yeager DS. Measurement matters: assessing personal qualities other than cognitive ability for educational purposes. *Educ Res*. 2015;44(4):237-251. doi:10.3102/0013189X15584327
- Patterson F, Zibarras L, Ashworth V. Situational judgement tests in medical education and training: research, theory and practice: AMEE Guide No. 100. *Med Teach*. 2016;38(1):3-17. doi:10.3109/0142159X.2015.1072619
- Webster ES, Paton LW, Crampton PES, Tiffin PA. Situational judgement test validity for selection: a systematic review and meta-analysis. *Med Educ*. 2020;54(10):888-902. doi:10.1111/medu.14201
- Dunleavy DM, Fraser RL. Predicting Student Professionalism in Medical School: A Situational Judgment Test: Presentation to the AAMC Annual Meeting. 2019. www.aamc.org/system/files/2021-03/services-sjt-lsl-update-2019-03012021.pdf
- Tiffin PA, Mwandigha LM, Paton LW, et al. Predictive validity of the UKCAT for medical school undergraduate performance: a national prospective cohort study. *BMC Med*. 2016;14(1):140 doi:10.1186/s12916-016-0682-7
- Patterson F, Cousins F, Edwards H, Rosselli A, Nicholson S, Wright B. The predictive validity of a text-based situational judgment test in undergraduate medical and dental school admissions. *Acad Med*. 2017;92(9):1250-1253. doi:10.1097/ACM.0000000000001630
- Dowell J, Cleland J, Fitzpatrick S, et al. The UK medical education database (UKMED) what is it? Why and how might you use it? *BMC Med Educ*. 2018;18(1):6. doi:10.1186/s12909-017-1115-9
- Greatrix R, Dowell J. UKCAT and medical student selection in the UK—what has changed since 2006? *BMC Med Educ*. 2020;20(1):292 doi:10.1186/s12909-020-02214-1
- Paton LW, Tiffin PA, Smith D, Dowell JS, Mwandigha LM. Predictors of fitness to practise declarations in UK medical undergraduates. *BMC Med Educ*. 2018;18(1):68. doi:10.1186/s12909-018-1167-5
- Sam AH, Bala L, Westacott RJ, Brown C. Is academic attainment or situational judgement test performance in medical school associated with the likelihood of disciplinary action? A national retrospective cohort study. *Acad Med*. 2021;96(10):1467-1475. doi:10.1097/ACM.0000000000004212
- Sahota GS, Taggar JS. The association between Situational Judgement Test (SJT) scores and professionalism concerns in undergraduate medical education. *Med Teach*. 2020;42(8):937-943. doi:10.1080/0142159X.2020.1772466
- General Medical Council. *Process for Completing UKMED Research*. The General Medical Council. October 8, 2019. https://www.ukmed.ac.uk/documents/UKMED_research_process.pdf
- University of Dundee. *HIC Trusted Research Environment*. University of Dundee. February 15, 2022. <https://www.dundee.ac.uk/hic/hic-trusted-research-environment/>
- Higher Education Statistics Agency. *Rounding and Suppression to Anonymise Statistics*. HESA. 2021. Accessed December 10, 2021. <https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics>
- Tiffin PA, Dowell JS, McLachlan JC. Widening access to UK medical education for under-represented socioeconomic groups: modelling the impact of the UKCAT in the 2009 cohort. *Bmj*. 2012;344:e1805. doi:10.1136/bmj.e1805
- Mwandigha LM, Tiffin PA, Paton LW, Kasim AS, Böhnke JR. What is the effect of secondary (high) schooling on subsequent medical school performance? A national, UK-based, cohort study. *BMJ Open*. 2018;8(5):e020291. doi:10.1136/bmjopen-2017-020291
- Smith D, Twigg E, Gurusamy G, Tang P. UKMED - 2020 V2 - October Data Dictionary https://www.ukmed.ac.uk/documents/UKMED_data_dictionary.pdf
- Patterson F, Martin S, Baron H, Fung K, Flaxman C. UKCAT Situational Judgement Test: Technical Report. 2014.
- McManus I, Dewberry C, Nicholson S, Dowell J. The UKCAT-12 study: educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a collaborative study of twelve UK medical schools. *BMC Med*. 2013;11(1):244. doi:10.1186/1741-7015-11-244
- Stata Statistical Software: Release 15*. StataCorp; 2017.
- 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;17:246.
- Chen H, Cohen P, Chen S. How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Communications in Statistics - Simulation and Computation*. 2010;39(4):860-864. doi:10.1080/03610911003650383
- Mwandigha LM. *Evaluating and Extending Statistical Methods for Estimating the Construct-Level Predictive Validity of Selection Tests*. University of York; 2017.
- Dore KL, Reiter HI, Kreuger S, Norman GR. Correction to: CASPer, an online pre-interview screen for personal/professional characteristics: prediction of national licensure scores. *Adv Health Sci*. 2017;22(5):1321-1322. doi:10.1007/s10459-017-9798-6
- Paton LW, McManus IC, Cheung KYF, Smith DT, Tiffin PA. Can achievement at medical admission tests predict future performance in postgraduate clinical assessments? A UK-based national cohort study. *BMJ Open*. 2022;12(2):e056129.
- Tiffin PA, Carter M. Understanding the measurement model of the UKCAT Situational Judgment Test: Summary Report. 2015. http://www.ukcat.ac.uk/App_Media/uploads/pdf/Understanding%20the%20measurement%20model%20of%20the%20UKCAT%20SJT.pdf
- Fonteyne L, Duyck W, De Fruyt F. Program-specific prediction of academic achievement on the basis of cognitive and non-cognitive factors. *Learn Individ Differ*. 2017;56:34-48. doi:10.1016/j.lindif.2017.05.003

34. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14(3):300-306. doi:10.1097/01.EDE.0000042804.12056.6C
35. Thorndike RL, Stein S. An evaluation of the attempts to measure social intelligence. *Psychol Bull*. 1937;34(5):275-284. doi:10.1037/h0053850
36. The Medical Schools Council. Entry requirements [For Medicine]. The Medical Schools Council. November 23, 2021. Accessed November 23, 2021. <https://www.medschools.ac.uk/studying-medicine/making-an-application/entry-requirements>
37. Muthen LK, Muthen BO, Asparouhov T. Chapter 10: Missing data. In: *Regression and Mediation Analysis Using MPLUS*. Muthen & Muthen; 2016:443-445.
38. Aldridge J, Bray S, David T. Medical student fitness to practise committees at UK medical schools. *BMC Res Notes*. 2009;2(1):97. doi:10.1186/1756-0500-2-97
39. Maher BM, Hynes H, Sweeney C, et al. Medical school attrition—beyond the statistics a ten year retrospective study. *BMC Med Educ*. 2013;13(1):13. doi:10.1186/1472-6920-13-13
40. De Corte W, Sackett P, Lievens F. Designing pareto-optimal selection systems: formalizing the decisions required for selection system development. *J Appl Psychol*. 2011;96(5):907-926. doi:10.1037/a0023298
41. 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:10.1111/medu.14606
42. Lievens F, Patterson F, Corstjens J, Martin S, Nicholson S. Widening access in selection using situational judgement tests: evidence from the UKCAT. *Med Educ*. 2016;50(6):624-636. doi:10.1111/medu.13060
43. 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:10.1111/medu.13605
44. Cipresso P. Modeling behavior dynamics using computational psychometrics within virtual worlds. *Front Psychol*. 2015;6:1725. doi:10.3389/fpsyg.2015.01725

How to cite this article: Tiffin PA, Sanger E, Smith DT, Troughton A, Paton LW. Situational judgement test performance and subsequent misconduct in medical students. *Med Educ*. 2022;1-10. doi:10.1111/medu.14801