

Widening access in selection using situational judgement tests: evidence from the UKCAT

Filip Lievens,¹ Fiona Patterson,² Jan Corstjens,¹ Stuart Martin³ & Sandra Nicholson⁴

CONTEXT Widening access promotes student diversity and the appropriate representation of all demographic groups. This study aims to examine diversity-related benefits of the use of situational judgement tests (SJTs) in the UK Clinical Aptitude Test (UKCAT) in terms of three demographic variables: (i) socio-economic status (SES); (ii) ethnicity, and (iii) gender.

METHODS Outcomes in medical and dental school applicant cohorts for the years 2012 ($n = 15\,581$) and 2013 ($n = 15\,454$) were studied. Applicants' scores on cognitive tests and an SJT were linked to SES (parents' occupational status), ethnicity (White versus Black and other minority ethnic candidates), and gender.

RESULTS Firstly, the effect size for SES was lower for the SJT ($d = 0.13$ – 0.20 in favour of the higher SES group) than it was for the cognitive tests ($d = 0.38$ – 0.35). Secondly, effect sizes for ethnicity of the SJT and cognitive

tests were similar ($d = \sim 0.50$ in favour of White candidates). Thirdly, males outperformed females on cognitive tests, whereas the reverse was true for SJTs. When equal weight was given to the SJT and the cognitive tests in the admission decision and when the selection ratio was stringent, simulated scenarios showed that using an SJT in addition to cognitive tests might enable admissions boards to select more students from lower SES backgrounds and more female students.

CONCLUSIONS The SJT has the potential to appropriately complement cognitive tests in the selection of doctors and dentists. It may also put candidates of lower SES backgrounds at less of a disadvantage and may potentially diversify the student intake. However, use of the SJT applied in this study did not diminish the role of ethnicity. Future research should examine these findings with other SJTs and other tests internationally and scrutinise the causes underlying the role of ethnicity.

Medical Education 2016; 50: 624–636
doi: 10.1111/medu.13060

Discuss ideas arising from the article at
www.mededuc.com/discuss.



¹Department of Personnel Management and Work and Organisational Psychology, Ghent University, Ghent, Belgium

²Department of Psychology, School of the Biological Sciences, University of Cambridge, Cambridge, UK

³Work Psychology Group, Derby, UK

⁴Institute of Health Sciences Education, Barts and The London, Queen Mary University of London, UK

Correspondence: Filip Lievens, Department of Personnel Management and Work and Organisational Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium. Tel: 00 32 9 264 64 53; Fax: 00 32 9 264 64 94; E-mail: filip.lievens@ugent.be

 INTRODUCTION

Widening access promotes student diversity and the appropriate representation of all demographic groups. Widening access to medical and dental education for a diverse group of students is important for various reasons. Interactions with a diverse group of peers during education and training enable students to develop their cultural competence and enhance their educational experiences.¹ At the broader societal level, a more diverse group of students and clinicians might make the future workforce more population-representative. In turn, this representation increases potential doctor–patient concordance in terms of background, for which evidence indicates improved patient satisfaction,^{2,3} access to medical care for minority populations⁴ and, in some cases, health outcomes.⁵

Diversity comprises many aspects.⁶ As well as ethnicity and gender, an important aspect of increasing diversity consists of widening access to health care education for students from all socio-economic backgrounds.⁷ Socio-economic status (SES) can be defined as the social standing or class of an individual or group. It is important because people from backgrounds of lower SES may be disadvantaged in achieving upward social mobility due to a lower likelihood of applying for higher education and participating in intellectually stimulating social networks.^{8,9} This study focuses on these three factors (SES, ethnicity and gender) in the context of diversity and widening access in medical and dental admissions.

Historically, widening access in medical and dental education has been hampered by the predominant use of academic and cognitively oriented assessments as formal admission tools: a wealth of evidence indicates that these measures tend to favour ‘traditional’ applicants (i.e. White, male and high SES applicants).^{10,11} Recent research has explored the use of aptitude tests in relation to diversity issues in medical school admissions. However, results so far have not shown significant benefits in terms of increasing diversity. For example, the UK Clinical Aptitude Test (UKCAT) seems to have minimal impact on widening access because it favours more ‘traditional’ applicants to medicine and dentistry: ‘...the UKCAT has an inherent favourable bias to male applicants and those from a higher socio-economic class’,¹² and ‘...most of the socio-demographic factors that predict A-level attainment

also predict UKCAT performance albeit less so’¹³ (see also^{14,15}). In the USA, similar findings regarding increasing representation of minority groups and admission scores have been reported. For instance, prior research found a substantial relationship between SES and cognitive ability test scores (estimated corrected mean correlation of 0.42 in the entire test-taking population)¹⁶ and identified that people from backgrounds of lower SES have more difficulties in entering and progressing through medical training.¹⁷

These results from the USA and UK confirm that other strategies¹⁸ to increase access to medical education for students from under-represented minority backgrounds should be sought and examined. The rationale for such a change refers to increasing acceptance of the proposal that selection to the health care professions should not be dependent on academic or cognitive assessments alone and that complementary means of assessing personal attributes deemed essential to health care providers are required.^{6,19,20} One of these strategies consists of complementing the traditional markers of academic attainment and cognitive ability tests with assessments of (inter)personal (i.e. non-academic) characteristics (e.g. integrity, adaptability) during medical and dental admissions examinations.

In addition to other selection methods, such as the multiple mini-interview (MMI),^{21,22} the situational judgement test (SJT) has emerged as a reliable instrument for assessing such non-academic constructs as complements to measures of cognitive ability. Situational judgement tests are complementary to MMIs and are often used in the earlier stages of selection because they can be computer-delivered and machine-marked and thus are suitable for screening high volumes of applicants efficiently. In SJTs, applicants are presented with scenarios in a written or video-based format and a list of several response alternatives from which they are asked to choose the options they would be likely to carry out in response to the given situation.^{23,24} Previous research in medical selection has shown SJTs to offer substantial predictive validity for non-academic attributes, such as empathy, integrity and teamwork skills (e.g.^{18,25}). Specifically, prior research has shown that SJTs can provide a valid way of assessing procedural awareness in terms of important (inter)personal characteristics among large groups of prospective students in medical admission examinations.^{26,27} This research identifies that students’ performance on SJTs predicts subsequent

performance in communication courses and clinical internships, and even as physicians 9 years later.^{26,27}

Yet, little research to date has explored the impact on widening access of selection methods that focus solely on non-academic attributes. Although SJTs have been linked to diversity in employment settings,^{28,29} there is no evidence to support the assumption that SJTs may also increase diversity in successful medical and dental school applicants.³⁰ As a result, a key research question has remained unanswered: What is the link between SJT scores and SES, ethnicity and gender in medical and dental admissions that use SJTs? To the present authors' knowledge, this is the first study to report the impact of SJTs in relation to widening access.

Therefore, this study's aim is three-fold. We examine the potential diversity benefits of SJTs in terms of: (i) SES; (ii) ethnicity, and (iii) gender. The context was the UKCAT, which recently incorporated an SJT alongside its existing cognitive ability assessments. Our expectation was that the roles of SES, ethnicity and gender in SJT performance would be low (by contrast with the roles of SES, ethnicity and gender in cognitive ability test scores) because SJTs are specifically designed to measure (inter)personal characteristics.^{10,23,31}

METHODS

Context

The data for this study were obtained from the UKCAT Consortium. The UKCAT is a computer-based standardised examination, administered in a fully proctored environment at test centres worldwide. It is taken by applicants to the undergraduate medical and dental education programmes offered by most UK universities. It was introduced in 2006 as a medical and dental school admission test with a three-fold purpose: firstly, to enhance the differentiation and fairness of selection to such highly oversubscribed programmes; secondly, to assess aptitudes that were considered to be vitally important for a successful medical or dental career, and, finally, to provide a test that would increase the potential participation of under-represented socio-economic groups. As detailed below, the UKCAT includes a battery of four cognitive ability tests. Since 2013 (pilot-tested in 2012), it has also encompassed an SJT for measuring interpersonal characteristics.

Sample

Data were derived from two cohorts. In 2012, a first cohort of 24 359 applicants (43.7% male; mean age: 19.39 years) took the UKCAT with the SJT included for piloting purposes. In 2013, a second cohort of 25 680 applicants (42.0% male; mean age: 19.15 years) completed a UKCAT of which the SJT represented an integral part.

Table 1 presents a detailed breakdown of the two cohorts in terms of age, gender, ethnicity, SES and educational background. The sample sizes and percentages in this table differ from those cited above because they are based on individuals for whom data on all study variables were available. Candidates with missing data on any of these variables (e.g. variables related to widening access) were excluded from this table and from our analyses, leaving cohorts numbering 15 581 for 2012 and 15 454 for 2013. We conducted missing data analyses to examine the impact of removing these data.

Measures

Situational judgement test

The written SJT was specifically designed to assess several (inter)personal constructs that are relevant in a potential future clinician, namely: integrity; perspective taking, and team involvement. Items were developed by trained organisational psychologists in collaboration with 27 clinician subject matter experts (SMEs). Of the SMEs, 37% were female and 11% were Black or belonged to minority ethnic groups (BME). They included senior lecturers, clinical tutors, heads of departments, directors of undergraduate education and consultants from a range of medical and dental schools throughout the UK. The SMEs provided item material and initial keys during telephone interviews with the organisational psychologists. Items were then reviewed in focus groups with those highly familiar with the role of a medical or dental student to ensure their realism and relevance to the target role. Eight focus groups were held across four locations and comprised role incumbents, lecturers and educators. The group was approximately 60% female and 28% BME. Items and keys were modified as required. Following this, a different group of 20 SMEs completed the items in test conditions. This group was approximately 40% female and 10% BME. Ten SMEs each completed half of the total number of items. The SMEs were instructed to rate either the appropriateness or

Table 1 Demographic data for the study groups in the cohorts of 2012 and 2013

Variable	Cohort of 2012 Study group (n = 15 581)	Cohort of 2013 Study group (n = 15 454)
Age group, n (%)		
< 16 years	7 (< 0.1%)	13 (< 0.1%)
16–19 years	11 017 (70.7%)	10 651 (68.9%)
20–24 years	3372 (21.6%)	3606 (23.3%)
25–34 years	1028 (6.6%)	994 (6.4%)
≥35 years	157 (1.0%)	190 (1.2%)
Gender, n (%)		
Male	6902 (44.3%)	6655 (43.1%)
Female	8679 (55.7%)	8799 (56.9%)
Ethnic group, n (%)		
White	9423 (60.5%)	9154 (59.2%)
Asian	4424 (28.4%)	4379 (28.3%)
Black	958 (6.1%)	1072 (6.9%)
Mixed race	776 (5.0%)	848 (5.5%)
Other	0	1 (< 0.01%)
NS-SEC (parents), n (%)		
Managerial and professional occupations	11 966 (76.8%)	11 756 (76.1%)
Intermediate occupations	625 (4.0%)	629 (4.1%)
Small employers and own account workers	1591 (10.2%)	1613 (10.4%)
Lower supervisory and technical occupations	660 (4.2%)	646 (4.2%)
Semi-routine and routine occupations	739 (4.7%)	810 (5.2%)
Highest level of qualification (candidate), n (%)		
No formal qualification	1435 (9.2%)	1439 (9.3%)
Below honours degree	9871 (63.4%)	9443 (61.1%)
Honours degree or above	4275 (27.4%)	4572 (29.6%)

NS-SEC = National Statistics Socio-Economic Classification. Total populations of test takers equalled 24 359 for 2012 and 25 680 in 2013. Candidates with missing data on any of the study variables were excluded from further analyses.

the importance of each of the response options. The rating scales ranged from 1 (very appropriate/important) to 4 (very inappropriate/not at all important). Items were piloted when there was a sufficient degree of agreement amongst SMEs with respect to the scoring key and the extent to which the item content was deemed appropriate and relevant. Sufficient agreement was ascertained by two criteria: (i) that a minimum of 70% of participants were in agreement with respect to whether or not the response option was broadly appropriate/important, and (ii) that a minimum of 40% of participants gave exactly the same response.

The 2013 SJT consisted of 66 operational items associated with 16 scenarios, for each test form. The 2012 pilot SJT consisted of between 63 and 68 items associated with 13 scenarios, for each test form. All scenarios were set in either a health care (medical, dental) or educational context. For each scenario, respondents were asked to rate between three and six response options. Response instructions referred to the rating of either the appropriateness or the importance of each of the response options. The rating scales ranged from 1 (very appropriate/important) to 4 (very inappropriate/not at all important). Responses were graded on a scale of 1–4 using a scoring convention that referred to how closely the test-taker's response matched the scoring key. Pilot data were used as additional evidence to validate SME keys. Item keys were modified in cases in which a highly performing candidate's choice of key differed from the existing key, providing that SME agreement criteria were still met.

In 2012, there were 18 different SJT forms. In 2013, there were six different SJT forms. Test takers were randomly assigned to a form, with the exception of those who were awarded extra time, all of whom took Form 1 in 2013. The use of different SJT forms is recommended to control item exposure and mitigate possible retest and coaching effects.^{25,27} Each form was constructed by selecting items from the total pool of available items. Forms were constructed to ensure that they contained a range of the three target domains (integrity, perspective taking, team involvement) and a varied spread of content themes.

The 2013 form construction used the 2012 item pilot data to equate the difficulty of the six operational test forms using an item response theory

(IRT) pre-equate approach.³² The authors can be contacted for more details about this procedure.

The reliability of internal consistency (Cronbach's alpha) of the SJT scores across the six 2013 operational forms ranged from 0.77 to 0.80. For scores on the 18 pilot SJT forms in 2012, internal consistency reliability ranged from 0.56 to 0.71. As in 2012 these were pilot forms only, reliability was corrected using the Spearman–Brown formula by including only those items with reasonable item partials (≥ 0.13) that may have been included in an operational test at this time. This then enabled an estimated reliability for each form of a 70-item test with similar quality items. These reliabilities ranged between 0.77 and 0.87.

Cognitive ability tests

The cognitive ability tests measured four mental abilities (verbal reasoning, quantitative reasoning, abstract reasoning and decision analysis) deemed to be important for medical and dental education. Verbal and quantitative reasoning measured candidates' ability to critically evaluate information that is presented in either a written or a numerical form. The abstract reasoning sub-test assessed candidates' ability to recognise relationships from information using convergent and divergent thinking. Finally, the decision analysis test assessed candidates' ability to make sound decisions and judgements based on complex information. For this study, a composite of the four cognitive ability tests was used for the analyses because the inter-correlations among the tests were substantial (2012, range: 0.30–0.50; 2013, range: 0.28–0.48) and each sub-test was designed to measure a different aspect of general mental ability.

Ethnicity

As part of the UKCAT, candidates are asked to provide information on their ethnicity. This information was used to create different ethnic groups (White, Asian, Chinese, Black, Mixed Race, Other and Not Declared). In a manner similar to that of earlier studies on widening access,^{12,15} we then recoded these categories to create a dichotomous variable representing White or BME groups. Specifically, all members of non-White communities (Asian, Chinese, Black, Mixed Race, Other) in the UK were included in the BME group. At the time of taking the UKCAT, BME members included in the present study were living in the UK and took the test in a certified test centre in

the UK. Worldwide, applicants living outside the UK can also sit the UKCAT at test centres in their country of residence. However, these non-UK applicants were excluded from the present study because details on their ethnicity were not available. For the purposes of further and more finely grained analyses, we created three more binary variables representing White or Black, White or Asian, and White or Mixed Race ethnic groups.

Socio-economic status

In student samples, SES is often operationalised by looking at a combination of three measures: parental education; parental income, and parental occupational status.¹⁶ Consistent with previous research in medical school admissions, in this study SES was operationalised by obtaining information about the candidates' parents' occupational status. Specifically, candidates were asked about their parents' or caregivers' occupations (types of job), employment status (unemployed, self-employed, employer, employee), supervisory status, and the size of the organisation for which they worked. The responses to these questions were then converted to the categories of the National Statistics Socio-Economic Classification (NS-SEC),³³ resulting in five distinct categories: managerial and professional; intermediate; small employers and own account workers; lower supervisory and technical, and semi-routine and routine occupations. As in past research,¹² we further collapsed these divisions into two categories (Higher SES and Lower SES). Higher SES included the occupational category 'managerial and professional occupations' (2012: $n = 11\,966$, 76.8%; 2013: $n = 11\,756$, 76.1%). Conversely, Lower SES consisted of all other categories (2012: $n = 3615$, 23.2%; 2013: $n = 3698$, 23.9%).

Analyses

Data were first subjected to a three-way analysis of variance (ANOVA) with factors SES (higher SES versus lower SES), Ethnicity (White versus BME), and Gender (male versus female) as independent variables and SJT scores as the dependent variable to investigate the relationships of these demographic variables with SJT performance and to check for potential interactions among them. We then added as covariate the composite score of the four cognitive ability tests and ran an analysis of covariance (ANCOVA) to check how much variance in SJT scores would still be attributable to our three factors after accounting for cognitive ability.

Next, we conducted effect size analyses (Cohen's d)³⁴ for the links between SES, ethnicity and gender, and SJT performance in both cohorts. In these analyses, we used the average SJT score across forms. Detailed results per form are available from ourselves. Cohen's d was obtained by taking the difference between the means of the two groups of interest, and dividing the result obtained by the pooled standard deviation (SD) of the two groups. On the basis of recommendations in previous research, we considered effect sizes of $d < 0.20$ as indicating little or no effect, $d \geq 0.20$ as representing a small effect, $d \geq 0.50$ as indicating a moderate effect, and $d \geq 0.80$ as demonstrating a large effect.^{34,35}

Finally, we also simulated various scenarios to examine the effect of using SJTs on the demographic profile (SES, gender and ethnicity) of admitted students. In these scenarios, the role of the SJT in the admittance policy changed.³⁶

RESULTS

Overall analyses

In the 2012 cohort, the three-way ANOVA revealed significant main effects for SES ($F_{(1,15573)} = 9.216$, $p < 0.01$, $\eta^2 = 0.01$), Ethnicity ($F_{(1,15573)} = 431.380$, $p < 0.001$, $\eta^2 = 0.03$) and Gender ($F_{(1,15573)} = 115.337$, $p < 0.001$, $\eta^2 = 0.01$) on SJT performance. In the 2013 cohort, the picture was similar: three main effects for SES ($F_{(1,15446)} = 28.561$, $p < 0.001$, $\eta^2 = 0.01$), Ethnicity ($F_{(1,15446)} = 779.838$, $p < 0.001$, $\eta^2 = 0.05$), and Gender ($F_{(1,15446)} = 111.818$, $p < 0.001$, $\eta^2 = 0.01$) emerged. In both cohorts, all interaction terms were non-significant. In the 2012 cohort, the results of the subsequent ANCOVAs showed a non-significant main effect of SES when cognitive ability was controlled for ($F_{(1,15572)} = 0.043$, $p = 0.835$). There were still main effects of Ethnicity ($F_{(1,15572)} = 261.958$, $p < 0.001$, $\eta^2 = 0.02$) and Gender ($F_{(1,15572)} = 157.934$, $p < 0.001$, $\eta^2 = 0.01$). In the 2013 cohort, ANCOVA results were similar, showing a non-significant main effect for SES ($F_{(1,15445)} = 0.353$, $p = 0.553$), and main effects for Ethnicity ($F_{(1,15445)} = 430.865$, $p < 0.001$, $\eta^2 = 0.02$) and Gender ($F_{(1,15445)} = 206.268$, $p < 0.001$, $\eta^2 = 0.01$). Again, none of the interaction terms were significant. Thus, in both cohorts, SES was no longer related to SJT scores when cognitive ability was controlled for.

Role of SES

Table 2 presents the means, SDs, significance tests and effect sizes of the cognitive ability test and SJT scores for the different SES groups. In the UKCAT 2012 cohort, we observed significant differences in cognitive ability test scores that favoured the higher SES group and gave $d = 0.38$, which is a small effect size. For the SJT scores, findings showed an average $d = 0.13$ in favour of the higher SES group, which is a little or no effect size. For the 2013 cohort, a similar pattern was observed. There were significant differences between the lower and higher SES groups on the cognitive ability tests composite score ($d = 0.35$). The average effect size associated with the SJT score was higher ($d = 0.20$) than in 2012, and can be classified as a small effect.

Role of ethnicity

Table 3 displays the means, SDs, significance tests and effect sizes of the cognitive ability tests and SJT scores for the two ethnicity groups. Results showed that the effect sizes obtained for the SJT and the cognitive ability tests were substantial and similar: White candidates scored on average half an SD higher than BME candidates on both the cognitive ability tests and the SJT. In 2012, the average d of the SJT scores was 0.47, which represents a small to moderate effect size, favouring White candidates, whereas it was 0.52 for the cognitive ability tests, which is a moderate effect size. In 2013, the average d of the SJT scores was 0.56, whereas that for the cognitive ability tests was 0.48.

We also conducted ancillary analyses in which we subdivided the group of BME candidates and compared them with White candidates. For the SJT, the largest effect sizes were found when White candidates were compared with Asian candidates ($d = 0.52$ and $d = 0.61$ in 2012 and 2013, respectively). Results for comparisons between White and Black candidates were $d = 0.43$ and $d = 0.55$ in 2012 and 2013, respectively. Comparisons between the White and Mixed Race groups gave the lowest mean score differences for the SJT ($d = 0.21$ and $d = 0.25$ in 2012 and 2013, respectively). Note that for the cognitive ability tests the largest differences in mean scores were observed when White candidates were compared with Black candidates ($d = 1.04$ and $d = 1.00$ for 2012 and 2013, respectively, representing large effect sizes). By contrast, effect sizes related to mean score differences between White and Asian candidates were less than half ($d = 0.45$

Table 2 Means, standard deviations (SDs), independent-samples t-tests and effect sizes of the cognitive ability test (CAT) and situational judgement test (SJT) scores according to socio-economic status (SES)

	High SES group		Low SES group		Mean difference	Cohen's <i>d</i>	95% CI of <i>d</i>
	<i>n</i>	Mean ± SD	<i>n</i>	Mean ± SD			
2012 cohort							
SJT	11 966	204.76 ± 11.70	3615	203.22 ± 12.14	1.54*	0.13	0.09–0.17
CAT	11 966	645.01 ± 64.51	3615	620.34 ± 66.66	24.67*	0.38	0.34–0.42
2013 cohort							
SJT	11 756	198.82 ± 15.54	3698	195.64 ± 17.41	3.19*	0.20	0.16–0.24
CAT	11 756	677.17 ± 72.16	3698	651.51 ± 76.48	25.66*	0.35	0.31–0.39

95% CI = 95% confidence interval.
**p* < 0.01.

Table 3 Means, standard deviations (SDs), independent-samples t-tests and effect sizes of the cognitive ability test (CAT) and situational judgement test (SJT) scores according to ethnicity

	White group		BME group		Mean difference	Cohen's <i>d</i>	95% CI of <i>d</i>
	<i>n</i>	Mean ± SD	<i>n</i>	Mean ± SD			
2012 cohort							
SJT	9423	206.55 ± 10.80	6158	201.11 ± 12.52	5.44*	0.47	0.44–0.51
CAT	9423	652.24 ± 59.45	6158	619.29 ± 70.11	32.95*	0.52	0.48–0.55
2013 cohort							
SJT	9154	201.61 ± 13.95	6300	192.94 ± 17.42	8.67*	0.56	0.53–0.59
CAT	9154	685.28 ± 66.31	6300	650.41 ± 79.49	34.87*	0.48	0.45–0.52

95% CI = 95% confidence interval; BME = black and minority ethnic groups.
**p* < 0.01.

and *d* = 0.42 for 2012 and 2013, respectively) and those recorded for the White and Mixed Race groups approximately a fifth (*d* = 0.19 and *d* = 0.18 for 2012 and 2013, respectively) of those observed in the comparison between White and Black candidates.

Role of gender

Table 4 presents means, SDs, significance tests and effect sizes of the cognitive ability test and SJT scores broken down by gender. Results showed a consistent pattern. Males had significantly higher scores than females on the cognitive ability tests (*d* = 0.17 and *d* = 0.16 in 2012 and 2013, respec-

tively, indicating little or no effect), whereas females significantly outperformed males on the SJT (*d* = -0.26 and *d* = -0.21 in 2012 and 2013, respectively, indicating small effect sizes). This confirms previous meta-analytic findings attesting to the higher scores obtained by female candidates on SJTs (average meta-analytic *d* = 0.11 in favour of women).

Missing data analyses

With respect to the full sample of applicants taking the UKCAT, both the 2012 and 2013 cohorts contained a substantial portion of missing values for the variables ethnicity (23.0% and 23.2% for 2012

Table 4 Means, standard deviations (SDs), independent-samples t-tests and effect sizes of the cognitive ability test (CAT) and situational judgement test (SJT) scores according to gender

	Males		Females		Mean difference	Cohen's <i>d</i>	95% CI of <i>d</i>
	<i>n</i>	Mean ± SD	<i>n</i>	Mean ± SD			
2012 cohort							
SJT	6902	202.71 ± 12.16	8679	205.76 ± 11.36	− 3.05*	− 0.26	− 0.23 to 0.29
CAT	6902	645.40 ± 65.38	8679	634.36 ± 65.83	11.04*	− 0.17	− 0.14 to − 0.20
2013 cohort							
SJT	6655	196.17 ± 16.51	8799	199.48 ± 15.57	− 3.31*	− 0.21	− 0.24 to − 0.18
CAT	6655	677.84 ± 75.31	8799	665.91 ± 72.59	11.93*	0.16	0.13 to 0.19

95% CI = 95% confidence interval.
**p* < 0.01.

and 2013, respectively), SES (12.7% and 12.0% for 2012 and 2013, respectively), and highest academic qualification (9.3% and 13.9% for 2012 and 2013, respectively). If applicants from BME backgrounds, low SES and with no formal education were to have systematically omitted an answer on the demographic questions, missing data may not be random and participants with these characteristics may be under-represented in the datasets. The reported effect sizes in the present study are then likely to represent underestimations. To test this assumption, we conducted multiple data imputations as a kind of sensitivity analysis in both the 2012 and 2013 datasets.^{13,37,38} The missing data imputation technique creates a predicted value for every missing data point based on all other data points present in the dataset. This predicted value is pooled from several simulated iterations (in the present study we ran 10 iterations). The analyses on the imputed dataset revealed the effect sizes to be similar to the effect sizes obtained with the missing values excluded, indicating that data were missing at random.

Use of SJTs and simulations about the demographic profile of admitted students

Although prior analyses examined the links between test scores and SES, ethnicity and gender, they do not show what might happen to the demographic composition of the student pool when SJTs are used to support admissions decisions. Given that we do not have detailed information on the policies of university admission boards, we ran different hypothetical scenarios³⁰ to simulate the effects of different admis-

sion policies on the numbers of students selected according to the three demographic groups.

The scenarios varied in terms of two key factors. Firstly, regarding the weighting of cognitive test and SJT scores, we distinguished among three scenarios: (i) a composite in which 100% weight was given to cognitive test scores and 0% weight was given to SJT scores (this was the situation in the past); (ii) a composite in which 80% weight was given to cognitive test scores and 20% to SJT scores (in this scenario each of the five tests was given the same weight), and (iii) a composite in which the weight was equally divided between cognitive test scores (50%) and SJT scores (50%) in the admission decision. Secondly, we distinguished among three selection ratios. That is, we manipulated our scenarios to reflect a university admissions board that: (i) selected only the top 15% of applicants (*stringent policy*); (ii) selected 50% of applicants, or (iii) rejected only the worst-scoring 15% of applicants (*lenient policy*). Cross-referencing these two factors led to nine possible simulated scenarios.

Results for the composition of selected students in terms of SES and gender for these nine simulated scenarios are presented in Table 5 (for SES) and Table 6 (for gender). Given that the simulated scenarios showed the use of the SJT to have little benefit in terms of diversifying for ethnicity, we omitted these analyses for ethnicity. Overall, the results confirm the effect size analyses described above. That is, using an SJT in addition to cognitive ability tests might enable admissions boards to select more students from lower SES groups. This would also lead to the selection of

Table 5 Simulations of class composition according to socio-economic status (SES) under three ratios of selectivity using different combinations of cognitive ability test (CAT) and situational judgement test (SJT) scores

	High SES			Lower SES		
	CAT	CAT + SJT		CAT	CAT + SJT	
		4 : 1	1 : 1		4 : 1	1 : 1
2012						
Stringent	64.9%	64.9%	62.9%	35.1%	35.1%	37.1%
Moderate	58.0%	58.1%	56.9%	42.0%	41.9%	43.1%
Lenient	52.8%	52.9%	52.4%	47.2%	47.1%	47.6%
All		76.8%			33.2%	
2013						
Stringent	63.8%	64.2%	63.2%	36.2%	35.8%	36.8%
Moderate	57.3%	57.5%	57.0%	42.7%	42.5%	43.0%
Lenient	52.6%	52.6%	52.5%	47.4%	47.4%	47.5%
All		76.1%			33.9%	

Selection ratios: Stringent = top 15%; Moderate = top 50%; Lenient = top 85%. CAT represents a scenario with an exclusive reliance on a cognitive ability test score composite (verbal reasoning, quantitative reasoning, abstract reasoning, decision analysis) in the selection decision. CAT + SJT 4 : 1 represents a selection scenario using a combination of 80% CAT scores supplemented with 20% SJT scores in the selection decision. CAT + SJT 1 : 1 represents a selection scenario using an equal weighting of 50% CAT scores and 50% SJT scores in the selection decision. For example, starting from the upper left corner 64.9% of applicants with high SES and 35.1% of applicants with lower SES are selected under the stringent selection ratio of admitting only the top 15% of candidates and using the scenario in which only the CAT scores matter in the selection decision.

more female students. The effects are greater when the SJT has more weight in the decision composite (i.e. 50%) and the selection ratio is more stringent (i.e. selecting the top 15%). For example, Table 5 shows that in 2012 for a selection ratio of 15%, weighting the SJT for 50% in the decision composite might ensure that 37.1% of the admitted students are from backgrounds of lower SES, which is an increase of 2% in comparison with the use of cognitive ability tests alone. Similarly, Table 6 shows that in the stringent selection ratio scenario, when the SJT is given a weighting of 50%, about 9% more females (52.3%) will be selected than when only cognitive ability tests are used (43.4%). Stronger effects on the representation of candidates of lower SES and women might be expected if even more weight (e.g. more than 50%) were to be given to the SJT in the decision composite.

DISCUSSION

The potential for SJTs to offer added value in selecting for non-academic attributes in health care education and training is a topical research agenda. As the current political priority within

many countries is to widen access,⁵ it is critical that any additional selection method also addresses this issue. To our knowledge, this study is the first to investigate whether SES, ethnicity and gender explain the performance of applicants on a bespoke SJT for medical and dental school admissions.

One key finding of this study is that the addition of an SJT to the UKCAT shows promise in widening access and balancing out the effects of the traditional reliance on indicators of academic attainment because SJTs do not seem to disadvantage students coming from low SES groups. Such an effect would also see a potential increase in the number of female candidates selected. Conversely, the cognitive ability test scores showed a significant relationship with SES, with their effect size being moderate. Note, however, that the effect size related to cognitive ability test scores in the UK data ($d = 0.35-0.38$) was somewhat lower than those previously found in the USA ($r = 0.42$ or approximately $d = 0.93$)¹⁶ and than those related to traditional markers of academic attainment in the UK such as A-level results.^{12,13}

Table 6 Simulations of class composition for different genders under three ratios of selectivity using different combinations of cognitive ability test (CAT) and situational judgement test (SJT) scores

	Males			Females		
	CAT	CAT + SJT		CAT	CAT + SJT	
		4 : 1	1 : 1		4 : 1	1 : 1
2012						
Stringent	56.6%	53.8%	47.7%	43.4%	46.2%	52.3%
Moderate	53.4%	51.9%	48.8%	46.6%	48.1%	51.2%
Lenient	51.1%	50.7%	49.6%	48.9%	49.3%	50.4%
All		44.3%			55.7%	
2013						
Stringent	56.3%	53.9%	48.8%	43.7%	46.1%	51.2%
Moderate	53.1%	52.0%	49.4%	46.9%	48.0%	50.6%
Lenient	51.1%	50.7%	49.8%	48.9%	49.3%	50.2%
All		43.1%			56.9%	

Selection ratios: Stringent = top 15%; Moderate = top 50%; Lenient = top 85%. CAT represents a scenario with an exclusive reliance on a cognitive ability test score composite (verbal reasoning, quantitative reasoning, abstract reasoning, decision analysis) in the selection decision. CAT + SJT 4 : 1 represents a selection scenario using a combination of 80% CAT scores supplemented with 20% SJT scores in the selection decision. CAT + SJT 1 : 1 represents a selection scenario using an equal weighting of 50% CAT scores and 50% SJT scores in the selection decision.

Another key finding of this study was that the effect sizes associated with ethnicity obtained for the SJT and the cognitive ability tests were substantial and similar. Thus, the addition of an SJT does not seem to increase diversity in terms of ethnicity. In the UK, the gap in the scores between White and BME candidates (about $d = 0.50$) was similar for the cognitive ability test and SJT scores. This runs counter to meta-analytic research in the USA, in which SJTs were found to have lower Black–White differences than cognitive ability tests among employee samples,³⁵ although larger differences have been found for SJTs in applicant samples and when the SJT assessed more cognitively oriented competencies.³¹ Another possible reason for these discrepant results is that the minority groups in the USA and UK are different. In the UK, Black people represent 6–7% of the minority group and people of Asian and Indian ethnicity make up the majority of the group. Another reason why some ethnic groups score lower on the SJT may be that the SJT situations, response options and scoring key reflect the cultural patterns of the majority group.²⁷ Clearly, future international research is needed to examine this explanation and other reasons for the ethnicity effects found in this study. This is a challenging issue because recent

research⁹ found that Black–White differences in cognitive ability tests were already large at a very young age (i.e. after 54 months) and remained stable.

At a practical level, our results suggest that the inclusion of an SJT in medical and dental school admissions settings does not disadvantage students coming from low SES backgrounds, and provide favourable evidence for using an SJT within a selection test battery to supplement existing cognitively oriented selection procedures. This is especially true as prior research also shows that SJTs enable recruiters to broaden the characteristics measured, including, for example, the personal attributes thought to be necessary in health care professionals.

This study is not without limitations. Firstly, all results are based on a specific SJT administered in the UK. Although the inclusion of two cohorts evidences the robustness of the results, we recommend that future studies with other SJTs in other countries should test the generalisability of our results. We want to highlight the importance of including a diverse group of SMEs in the SJT development and scoring process, as we did in this study. Results may

be expected to differ when the SJT captures more cognitively oriented constructs or requires greater reading comprehension skills. In fact, additional analyses showed that when the correlation between SJT scores and cognitive ability test scores was higher, the SJT's effect sizes (for both SES and ethnicity) were also higher.²⁹ These results are available from the authors.

Secondly, this study looked only at a snapshot of the education pathway.³⁹ That is, we examined the effects of the use of an SJT on admissions test performance. Future longitudinal studies are therefore needed to examine whether the use of SJTs leads to: (i) the attaching of substantial weight to SJT scores by medical and dental schools for the purpose of making admission decisions; (ii) more offers for students from under-represented groups; (iii) better access for students from under-represented groups to medical and dental schools in comparison with other health care programmes, and (iv) successful progress and a decrease in the number of students from under-represented groups who drop out. Only when those questions are answered can more definitive conclusions about the value of SJTs for increasing diversity in medical and dental education be drawn.

Thirdly, there were some methodological limitations. We could not correct the effect sizes for unreliability because item-level data regarding the selection procedures (cognitive ability tests and SJTs) were not available. In addition, as a result of missing data we had to remove a substantial part of our sample from the analyses (see 'Missing data analyses'). Nevertheless, missing data imputation revealed that our results were robust against the list-wise deletion procedure deployed prior to our data analyses.

We suggest that future research investigates whether specific characteristics of SJTs moderate the effects of SJT performance on diversity. Such moderators might include the complexity of the language used in the SJT or the type of contextualisation of the SJT. For example, some organisations may situate SJT items in a health care context. For instance, in the SJTs used in the Flemish admission examination,²⁷ students are presented with basic interpersonal interactions between a physician and a patient (e.g. establishing rapport with the patient, requesting more information). Such a medical context might favour students with specific backgrounds (e.g. a parent or close family member who is a physician). Other

organisations might situate similar incidents in an educational context (e.g. establishing rapport with a new student). In this case, one might assume that these less contextualised situations do not favour students with specific backgrounds. Still other organisations may decide to use a mixture of the two item types (such as the UKCAT in this study). In the future, we intend to conduct a detailed item-level analysis to examine the relationship of these different types of contextualisation with background variables such as ethnicity and SES. In addition, it would be valuable to examine potential differences in terms of face validity, applicant perceptions and stakeholder acceptance between these differentially contextualised SJT items.

Contributors: FL formulated the original idea for the study and drafted the manuscript. FP contributed to the design of the study and of the measures, the collection, analysis and interpretation of data, the writing of initial drafts and the revision of the manuscript. JC conducted the data cleaning and all data analyses and wrote the Methods and Results sections of the manuscript. SM contributed to the original conception of the study, the design of the measures, the analysis and interpretation of data, and the revision of the manuscript. SN contributed to the design of the study and of the measures, the collection of data and the revision of the manuscript. All authors approved the final manuscript for publication and have agreed to be accountable for the accuracy and integrity of the work.

Acknowledgements: The authors gratefully acknowledge the UKCAT Board and Pearson VUE for access to the data. We also thank Fran Cousans for supporting the preparation of the final manuscript.

Funding: none.

Conflicts of interest: SN currently chairs the UKCAT Research Group and stepped down as chair of the UKCAT Consortium in 2014. She contributed to the conception of the study and writing of the manuscript but did not directly analyse the data or influence the interpretation of the data analysis.

Ethical approval: this work was carried out in accordance with the Declaration of Helsinki. The study was approved by Ghent University's (Department of Psychology) human subject committee.

REFERENCES

- Whitla DK, Orfield G, Silen W, Teperow C, Howard C, Reede J. Educational benefits of diversity in medical school: a survey of students. *Acad Med* 2003;**78** (5):460–6.
- Laveist TA, Nuru-Jeter A. Is doctor–patient race concordance associated with greater satisfaction with care? *J Health Soc Behav* 2002;**43** (3):296–306.

- 3 Paez KA, Allen JK, Carson KA, Cooper LA. Provider and clinic cultural competence in a primary care setting. *Soc Sci Med* 2008;**66** (5):1204–16.
- 4 Deas D, Pisano ED, Mainous AG III, Johnson NG, Singleton MH, Gordon L, Taylor W, Hazen-Martin D, Burnham WS, Reves JG. Improving diversity through strategic planning: a 10-year (2002–2012) experience at the medical university of South Carolina. *Acad Med* 2012;**87** (11):1548–55.
- 5 Cohen JJ, Steinecke A. Building a diverse physician workforce. *JAMA* 2006;**296** (9):1135–7.
- 6 Eva KW. Moving beyond childish notions of fair and equitable. *Med Educ* 2015;**49** (1):1–3.
- 7 Cleland JA, Nicholson S, Kelly N, Moffat M. Taking context seriously: explaining widening access policy enactments in UK medical schools. *Med Educ* 2015;**49** (1):25–35.
- 8 Sackett PR, Kuncel NR, Beatty AS, Rigdon JL, Shen W, Kiger TB. The role of socioeconomic status in SAT-grade relationships and in college admissions decisions. *Psychol Sci* 2012;**23** (9):1000–7.
- 9 Cottrell JM, Newman DA, Roisman GI. Explaining the Black-White gap in cognitive test scores: toward a theory of adverse impact. *J Appl Psychol* 2015;**100** (6):1713–36.
- 10 Woolf K, Potts HWW, McManus IC. Ethnicity and academic performance in UK trained doctors and medical students: systematic review and meta-analysis. *BMJ* 2011;**342**:d901.
- 11 Girotti JA, Park YS, Tekian A. Ensuring a fair and equitable selection of students to serve society's health care needs. *Med Educ* 2015;**49** (1):84–92.
- 12 James D, Yates J, Nicholson S. Comparison of A level and UKCAT performance in students applying to UK medical and dental schools in 2006: cohort study. *BMJ* 2010;**340**:c478.
- 13 Tiffin PA, McLachlan JC, Webster L, Nicholson S. Comparison of the sensitivity of the UKCAT and A levels to sociodemographic characteristics: a national study. *BMC Med Educ* 2014;**14** (7):1–12.
- 14 Mathers J, Sitch A, Marsh JL, Parry J. Widening access to medical education for under-represented socioeconomic groups: population based cross sectional analysis of UK data, 2002–6. *BMJ* 2011;**342**:d918.
- 15 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.
- 16 Sackett PR, Kuncel NR, Arneson JJ, Cooper SR, Waters SD. Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychol Bull* 2009;**135** (1):1–22.
- 17 Cohen JJ. Disparities in health care: an overview. *Acad Emerg Med* 2003;**10** (11):1155–60.
- 18 Patterson F, Knight A, Dowell J, Nicholson S, Cousans F, Cleland J. How effective are selection methods in medical education and training? Evidence from a systematic review. *Med Educ* 2016;**50** (1):36–60.
- 19 Francis R. *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry: Executive Summary*. London: Stationery Office 2013.
- 20 Patterson F, Prescott-Clements L, Zibarras L, Edwards H, Kerrin M, Cousans F. Recruiting for values in healthcare: a preliminary review of the evidence. *Adv Health Sci Educ Theory Pract* [internet]. 2015; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25616718>
- 21 Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Med Educ* 2004;**38** (3):314–26.
- 22 Eva KW, Reiter HI, Trinh K, Wasi P, Rosenfeld J, Norman GR. Predictive validity of the multiple mini-interview for selecting medical trainees. *Med Educ* 2009;**43** (8):767–75.
- 23 McDaniel MA, Hartman NS, Whetzel DL, Grubb WL. Situational judgement tests, response instructions, and validity: a meta-analysis. *Pers Psychol* 2007;**60** (1):63–91.
- 24 Motowidlo SJ, Dunnette MD, Carter GW. An alternative selection procedure: the low-fidelity simulation. *J Appl Psychol* 1990;**75** (6):640–7.
- 25 Patterson F, Ashworth V, Kerrin M, O'Neill P. Situational judgement tests represent a measurement method and can be designed to minimise coaching effects. *Med Educ* 2013;**47** (2):220–1.
- 26 Lievens F, Sackett PR. The validity of interpersonal skills assessment via situational judgement tests for predicting academic success and job performance. *J Appl Psychol* 2012;**97** (2):460–8.
- 27 Lievens F. Adjusting medical school admission: assessing interpersonal skills using situational judgement tests. *Med Educ* 2013;**47** (2):182–9.
- 28 Ployhart RE, Holtz BC. The diversity–validity dilemma: strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Pers Psychol* 2008;**61** (1):153–72.
- 29 Whetzel DL, McDaniel MA, Nguyen NT. Subgroup differences in situational judgement test performance: a meta-analysis. *Hum Perform* 2008;**21** (3):291–309.
- 30 Schmitt N, Keeney J, Oswald FL, Pleskac TJ, Billington AQ, Sinha R, Zorrie M. Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *J Appl Psychol* 2009;**94** (6):1479–97.
- 31 Roth PL, Bobko P, Buster MA. Situational judgement tests: the influence and importance of applicant status and targeted constructs on estimates of Black-White subgroup differences. *J Occup Organ Psychol* 2013;**86** (3):394–409.
- 32 Patterson F, Baron H, Ashworth V, Martin S. UKCAT situational judgement test IRT appendix to the UKCAT SJT 2012 technical report. Prepared for the UKCAT Test Development Committee and

- Pearson VUE. Derby: Work Psychology Group 2013.
- 33 Office for National Statistics. *National Statistics Socio-Economic Classification User Manual*. Newport: Office for National Statistics 2005.
- 34 Cohen J. A power primer. *Psychol Bull* 1992;**112** (1):155–9.
- 35 Bobko P, Roth PL. Reviewing, categorising, and analysing the literature on Black-White mean differences for predictors of job performance: verifying some perceptions and updating/correcting others. *Pers Psychol* 2013;**66** (1):91–126.
- 36 De Corte W, Lievens F. A practical procedure to estimate the quality and the adverse impact of single-stage selection decisions. *Int J Select Assess* 2003;**11** (1):87–95.
- 37 Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;**338**: b2393.
- 38 Newman DA. Missing data: five practical guidelines. *Organ Res Methods* 2014;**17** (4):372–411.
- 39 Lievens F. Diversity in medical school admission: insights from personnel recruitment and selection. *Med Educ* 2015;**49** (1):11–4.

Received 22 August 2015; editorial comments to author 19 October 2015, 26 January 2016; accepted for publication 26 February 2016