Patterson, F., Ashworth, V., Baron, H., Thomas, S., & Smith, D. (2017). *The contribution of "Value-add" methodology to understanding medical school outcomes*. General Medical Council.

Work Psychology Group
Thinking differently

University of BRISTOL

General Medical Council
Regulating doctors
Ensuring good medical practice

The contribution of "Value-add" methodology to understanding medical school outcomes

Authors

Work Psychology Group

Professor Fiona Patterson

Vicki Ashworth

Helen Baron

University of Bristol

Sally Thomas

GMC

Daniel Smith

March 2017

1. Executive Summary

1.1. The aim of the project was to look at the potential for applying a 'value-add' methodology to the exploration of performance outcomes of medical students to better understand the indicators of how different institutions progress students and to look at the impact for widening participation.

1.2. The research questions of this study are examples of questions that the methodology might be able to address namely:

1.2.1. Are there differences between medical schools in the performance on qualification of candidates of similar ability on admission (i.e. controlling only for prior attainment) – that is, are some schools better at progressing students at medical school?

1.2.2. Is student progress typically associated with the widening access status of the students (i.e. controlling for both prior attainment and student individual characteristics)?

1.3. While the methodology has limitations and interpretation of results for a particular school needs to take account of the statistical uncertainty and degree of measurement error, work in secondary education has suggested that prior attainment and other factors can explain around 75% of the between school variance and around 50% total variance in GCSE or equivalent outcomes (Thomas, 2001).

1.4. A clear model of the impact of prior attainment should improve the detection of the impact of demographic variables such as gender and BME status as well as socio-economic variables. Such results would support our understanding of factors impacting wider participation initiatives.

1.5. The data set consisted of the two cohorts which started their medical education at a UK medical school in 2007 or 2008, who were educated in the UK and did not have entry qualifications at degree level or above. The analysis is based on a data set collected by the Higher Education Statistics Agency Limited ("HESA") and provided to the GMC ("HESA Data"). 10,036 cases met the initial selection criteria.

1.6. Two kinds of measures are required for the study – measures of prior attainment which show the ability of students when they enter medical school and outcome measures which indicate the performance of students at the end of their studies. In addition demographic variables are needed to address the second hypothesis.

1.7. Two measures of prior attainment were used. The first was the HESA Tariff scores based on attainment at secondary schools – typically A-levels and Highers. A number of analyses were undertaken to consider the best way to process these scores. The standard HESA tariff values were used over the UKCAT 12 approach because they provided increased variance. Scores below 300 were dropped because they indicated incomplete scores or unlikely levels of qualification for medical school even in a widening access context. The second was scores on the United Kingdom Clinical Aptitude Test (UKCAT) for entry to medical school because it represented a different intake measure and was available for a high proportion of candidates.

1.8. Three outcome measures were considered. (1) Academic outcomes were represented through the Educational Performance Measure Decile (EPM). These scores could not be used as the method of calculation removes between school differences. (2) The ARCP is a performance rating from F1 and F2 clinical placements and represents clinical performance. Unfortunately the measurement scale is only designed to differentiate competent levels of performance from non-competent. The result is that 88-90% of individuals receive the highest grade meaning that it does not differentiate outcomes finely enough. (3) The situational judgement test (SJT) used as part of the application process to foundation training programmes which measures approach to medical practice through a number of contextual competencies identified as important. There is an element of procedural knowledge but it does not measure medical knowledge or clinical competence. The SJT was chosen because it was the variable with the greatest score variance while allowing comparisons between medical schools. For the available measures the correspondence between the cognitive/academic prior attainment measures and the contextual behaviour measured by the SJT is weak and ideally a comparison with a more academic outcome measure would have been desirable. However, given the key aim was to explore and illustrate value-added methodology for medical schools, this was considered an acceptable limitation.

1.9. A literature review was undertaken to identify target variables for the widening participation part of the study. Socio-economic variables identified included the socio-economic classification of the student's family based on parents' occupation (SEC), parental education, residence in an area of low participation in higher education, claiming free school meals and/or income support and attendance at independent or state school. Demographic variables including age, gender, ethnicity and disability status were also collected.

1.10. 6,978 cases had full data for the prior attainment and outcome measures and 4,691 had data for all the background variables as well.

1.11. The first model applied to the data is a baseline model identifying the extent to which the outcome measure differs across the 29 medical schools studied. The variance partition coefficient was 0.04 showing that 4% of the variance in SJT scores was attributable to medical school attended. Typically studies of secondary school performance show greater differences between schools (7-10%) but 4% is a statistically significant amount of variance with explanatory power.

1.12. When prior attainment was added to the model significantly more of the SJT variance was explained. The total, 6-7%, is not large which may reflect the difference between the focus of the prior attainment measures (cognitive) and that of the outcome measure (contextual behaviour) and/or the restricted nature of the prior attainment measures given competition/selection for entry to medical school. However prior attainment explained around 50% of the variation in scores between medical schools suggesting that differences in student intake accounts for the differences between schools as much as the teaching approach and other unmeasured school process factors not accounted for in the analysis. Nevertheless, after controlling for prior attainment, around a third of the schools showed significantly different value added mean outcomes from the sample mean.

1.13. When the background and socio-economic variables were added to the model a further 3% of total variance in student outcome scores was explained. The largest effects were found for gender and BME status. These accounted for between a fifth and a quarter of the explained variance in SJT scores with women and White students having higher average SJT scores. There was a small but statistically significant impact of age with younger candidates tending to have higher SJT scores. Disability status had no significant impact.

1.14. Of the socio-economic variables, only whether the student had received free school meals and/or whether the family received income support had a statistically significant impact on outcome when prior attainment and other back ground factors were taken into account. Students from families receiving benefits performed a little less well with an average difference of less than 0.1 of a standard deviation. Other SES variables such as parents' profession and education showed significant differences in the SJT scores before prior attainment and other back ground factors were taken into account, but in the multilevel model including prior attainment and FSM/income support, which is usually more sensitive to differences, they did not show significant effects.

1.15. These results show that taking into account prior attainment is important in understanding outcomes of medical schools even with the poor match between the variables in this study, they accounted for around 50% of the variance between schools. In addition, modelling the impact of prior attainment supported the significant impact of widening participation variables on student progress at medical school but suggested that the impact of parents' education and background could be accounted for by other demographic factors

1.16. Because of the poor match between the prior attainment and outcome variables these results need to be considered as provisional. Further studies using better matching between prior attainment and outcome variables and samples based on the year of graduation are recommended, including consecutive cohorts of students, so that more rigorous findings including time trends can be modelled over more than one year group.

Acknowledgements

The authors would like to acknowledge the support of the GMC and UKMED in producing this report.

This report uses data from the Higher Education Statistics Agency Limited

Source: HESA Student Record 2002/03 to 2014/15

Note on competing interests of authors

Fiona Patterson, Vicki Ashworth and Helen Baron provide advice on selection methodology to Health Education England and UKCAT through the Work Psychology Group. None of the authors receive royalties for medical schools' use of the UKCAT SJT or UKFPO's use of the SJT for FPAS.

Work Psychology Group
Thinking differently

Contents

2. Overview of the Project

The aim of the project was to look at the potential for applying a 'value-add' methodology to the exploration of performance outcomes of medical students to better understand the indicators of future performance, how different institutions progress students and to look at the impact for widening participation.

Rationale for Value-added Methodology

2.1. The value-added concept rests on the assumption that educational institutions, including both schools and Higher Education Institutions (HEIs), add 'value' to the achievement of their students. This approach uses statistical techniques (e.g. multilevel modelling) to produce an estimate of the **extra value** that is added by schools to student attainment **over and above** the progress or improvement that might normally be expected based on a measure of ability and attainment on entry to the school.

2.2. The methodology involves comparing different models to separate out the effect of the **school experience** on individual student outcomes (what students achieve) and the extent to which **student intake characteristics** affect student outcomes. Therefore, accurate baseline information about students' prior attainment is crucial to calculate the value-added component (Thomas & Mortimore, 1996, Thomas, 2001). Value-added measures thus seek to establish whether students in some schools/HEIs make relatively greater or less progress than those in other schools/HEIs over a specified period, such as from beginning to end of primary or secondary schooling or during HEI undergraduate programmes. The most effective schools/HEIs would be those where student progress exceeds expectation. These measures therefore provide both an indicator of a school's effectiveness and a tool for teaching staff to use to analyse the extent to which they have effectively raised student achievement.

2.3. In terms of application, similar to other quantitative measures, value-added measures are essentially **estimates** and there are some limitations to the methodology and approaches to school evaluation which need to be well understood (Goldstein & Thomas, 1996). These include the issues of measurement error and the need to always consider the uncertainty associated with estimating individual school's value-added scores (i.e. via the statistical confidence intervals). Moreover, given value-added techniques have so far rarely been applied in the context of HEIs, this study is exploratory in nature, seeking to establish whether value-added techniques are a valid method to evaluate the educational impact of different medical schools.

Key Outcome and Prior Attainment Measures in Medical Education

2.4. A number of measures are used to select candidates for university places to study medicine. For non-graduate entry courses, these include level 3 educational qualifications such as A-levels or Scottish Highers and aptitude test results such as the United Kingdom Clinical Aptitude Test (UKCAT) scores. Each medical school determines its own approach to selection within broad parameters. However typically there are some ten applicants for every place (UCAS Analysis and Research, 2016a, 2016b) so the process is highly competitive.

2.5. There are a variety of outcome measures. These vary from academic results (examination scores, Educational Performance Measure – deciles (Medical Schools Council, 2011) to reviews of progression in supervised practice (Annual Review of Competence Progression, ARCP). In addition, medical school graduates become applicants for more advanced training and undergo assessment for selection for places on different courses (applications to foundation and later specialty training). They may be assessed on additional measures such as tests of clinical knowledge, situational judgement tests (SJT), interviews and objective structured clinical examinations (OSCE). During speciality training junior doctors will sit membership examinations offered by the different Royal Colleges consisting of a mixture of written examinations and OSCEs.

2.6. This project focused on the prediction of performance in the **final year of medical school**. The predictors chosen were those available for the largest proportion of the cohort, namely academic qualifications calibrated according to the tariff system used by UCAS and provided by HESA ("HESA", 2016) and UKCAT scores ("UKCAT About the Test", 2016) which, while required by only about three-quarters of medical schools, is taken by most applicants since they apply to multiple schools and will typically need UKCAT scores for at least one of them.

2.7. A number of outcome variables were considered for use. For the purpose of evaluating value-add methodology it is desirable that there is a good relationship between predictors and outcomes. Degree classification is often used as an outcome measure in studies of higher education, however medical degrees are not classified beyond pass/fail. (A recent large scale analysis of degree outcomes by HEFEC (2014) excluded medical students for this reason). Section 3 below discusses the methodology of the study and the reasons for choosing the variables to be studied. The relationship with the SJT scores in the application process to the foundation programme were found to be the strongest and therefore this variable was considered the only adequate measure to be used as a training outcome variable for the purposes of this exploratory study.

Widening Participation

2.8. This study also seeks to explore the typical impact of **students' individual background characteristics** on their progress at medical school. These student background characteristics reflect some of the key factors that have been highlighted by the press and other organisations such as the Sutton Trust concerned with widening participation in Higher Education, especially in supporting better access to HEIs for disadvantaged students from low income families. Moreover, although prior attainment is the key explanatory factor to control for measuring institutional value-add, including additional student background factors as explanatory variables may also be beneficial in fine-tuning institutional value-added scores (depending on the evaluation purpose of the measures created).

2.9. Recent initiatives to widen access to medical school (Selecting for Excellence, Medical Schools Council, 2016) have aimed to increase the chances of able candidates from lower socio-economic backgrounds to attain places, to address their under-participation compared to candidates from more privileged backgrounds. There are different approaches to measuring socioeconomic background for different purposes and there are a number of indicators and correlates which have some validity in the context of educational achievement.[1] These include parents' education and employment, school attended, area of living and a history of receipt of benefits such as income support and free-school meals. Indicators of several of these were explored in the modelling.

2.10. Figure 1 shows the outline of the study design.

Research Questions

2.11. The research questions addressed to illustrate the value added methodology in this study are therefore:

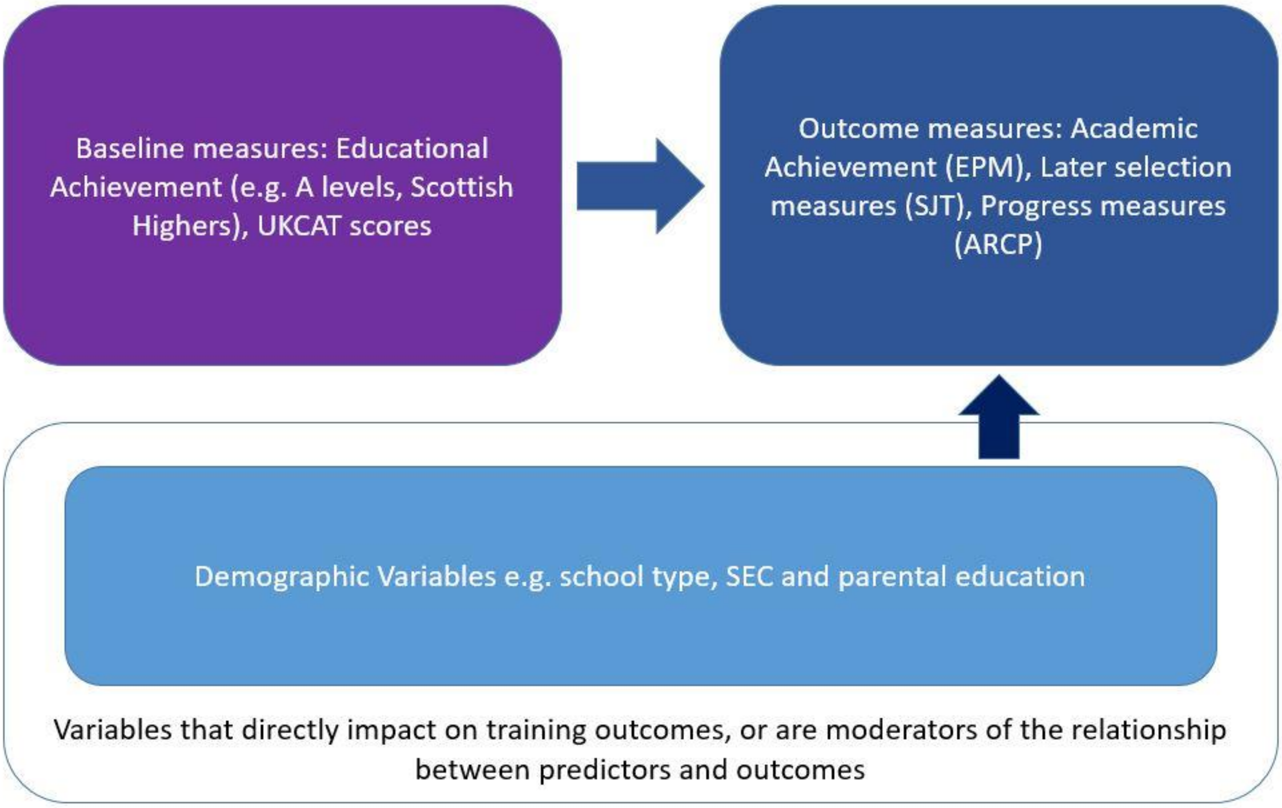
2.11.1. **Are there differences between medical schools in the performance on qualification of candidates of similar ability on admission (i.e. controlling only for prior attainment) – that is, are some schools better at progressing students at medical school?**

2.11.2. **Is student progress typically impacted by the widening access status of the students (i.e. controlling for both prior attainment and student individual characteristics)?**

---

[1] American Psychological Association. Education and Socioeconomic Status. https://www.apa.org/pi/ses/resources/publications/education.aspx

Figure 1: Outline of the use of demographic variables in the analysis of value-added by medical schools



3. Literature Review.

3.1. The first stage of the project was to identify variables relevant to widening access to include the modelling that might be expected to impact on outcomes. A literature review was undertaken to identify potential indicators of socio-economic position to use in the widening participation elements of the study. Information regarding the search strategy, including search terms, databases searched and inclusion/exclusion criteria can be found in Appendix A.

3.2. The review aimed to collate and summarise the published literature on demographic variables that may impact on medical students' performance at the point of entry to medical school, during their medical education, and consequently at the end of their undergraduate education.

3.3. Demographic variables identified in the review are likely to be highly confounded with other measures (Do, Parry, Mathers, & Richardson, 2006), and so caution should be exercised when considering the impact of variables in isolation. For example, it is highly unlikely that parental education, household income and postcode are mutually exclusive.

3.4. Although there are trends across all demographic variables identified in this review, there also remain notable discrepancies in the literature. That is: *no variable can be said to have a definite impact on attainment or success at the point of entry to medical school or during training*. Variables identified in this section as having the potential to impact on the relationship between students' performance at the point of entry to medical school and key performance outcomes are explored in the statistical analysis and modelling.

Socio-Economic Classification (SEC) and Parental Education

3.5. Putting SEC into context, the socio-economic difference in participation in higher education (HE) and achievement worsened in the UK during the 1980s and early 1990s (Blanden & Machin, 2004; Galindo-Rueda, Marcenaro-Gutierrez, & Vignoles, 2004; Glennerster et al., 2001; Machin & Vignoles, 2004). Whilst it appears to have narrowed since then (Raffe, 2007), with ethnic minorities and women no longer underrepresented in UK medical schools, lower socioeconomic groups still are (Secretary of State for Education and Skills, 2004).

3.6. With regards to identifying individuals in the lower socioeconomic groups, Do et al. (2006) cautioned against the sole use of parental occupation or area-based measures to monitor performance of students given the proportion of applicants who do not provide this information and the age and ethnic group differences in these groups. For example, Do and colleagues found that White applicants were more likely to be from professional/managerial classes (72%) than were Black (60%) or Asian (60%) applicants. They also note that the increased awareness of widening access policies has led to the perception among some applicants that some form of positive discrimination may be occurring in favour of "disadvantaged" groups, meaning that some applicants may "downgrade" their parents' jobs. At present, it is not possible to verify the factual accuracy of candidates' reported demographic data, although evidence of receipt of benefits may be required to claim bursaries. No published empirical data was found in the current review to support Do and colleagues' assertions, but there have been reports in recent years that university admissions procedures may 'favour the poor' (e.g. The Guardian, "University Admission 'Favours the Poor'", 2002).

3.7. Despite these possible concerns with classifying individuals into different socioeconomic groups, studies have generally found that an individual's probability of participating in HE is significantly determined by parental level of education and/or SEC (Blanden & Gregg, 2004; Carneiro & Heckman, 2002). Crawford (2014) offered a possible explanation for this finding, as she identified that large raw differences in university outcomes between individuals from different socio-economic backgrounds can largely be explained by the fact that they arrive at university with very different levels of human capital. Crawford defined human capital as students' existing knowledge and skills as gained by academic qualifications at Key Stages 2, 4 and 5, school type, and financial background measured as Index of Multiple Deprivation and ACORN type (constructed using information on socio-economic characteristics, financial holdings and property details; available for neighbourhoods of approximately 15 households). Once differences in human capital are accounted for, the differences in degree outcomes by socio-economic background become smaller, but remain significantly different from zero. Those from higher socio-economic backgrounds are 3.4% less likely to drop-out, 5.3% more likely to graduate and 3.7% more likely to graduate with a first or 2:1 degree classification than those from lower socio-economic backgrounds. It is notable that these findings are in stark contrast to similar analysis by school characteristics (Crawford, 2014; HEFCE, 2014) which show that, amongst students with the same grades on entry to university, those from worse-performing schools are less likely to dropout, more likely to complete their degree and more likely to obtain a first or 2.1 degree than those from better-performing schools (see 'School Type', below).

3.8. Nonetheless, Seyan et al (2004) found that in the year 2000, pupils from NS-SEC ("National Statistics Socio-Economic Classification", 2005) class 1 were around 100 times more likely to gain a place at medical school than those from classes 4 or 5 (for both White and Black students). Other longitudinal research suggests that although large differences in HE participation rates and participation rates at high status universities by socio-economic background do exist, these differences are substantially reduced once prior achievement is included (Chowdry, Crawford, Dearden, Goodman, & Vignoles, 2013). These differences in attainment at entry to HE come about largely because lower SES pupils do not achieve as highly in secondary school as their more advantaged counterparts, confirming the general trend in the literature that socio-economic differences emerge relatively early in individuals' lives. Moreover, these findings hold for both state and independent school pupils (see 'School Type', below).

3.9. Performance on medical school entry examinations such as aptitude tests may also be impacted by socio-economic status; contributing to lower proportions of students from lower SECs achieving a place at medical school. For example, James et al. (2010) used NS-SEC data collected from candidates sitting the UKCAT to derive categories of managerial/profession versus all other occupations. Those from the top professional/managerial backgrounds performed significantly better in all subtests of the UKCAT than did those from all other backgrounds ($p<0.001$ in all cases) (although note that this was before the inclusion of the UKCAT SJT, which shows minimal differences between candidates from different SECs, (Lievens, Patterson, Corstjens, Martin, & Nicholson, 2016). Similarly, Tiffin et al. (2014) found that candidates from non-professional socio-economic backgrounds were observed to achieve, on average, lower scores on both the UKCAT and at A-Level, even after controlling for the effects of other predictor variables (again this was before inclusion of the SJT).

3.10. Internationally, the pattern of results is less clear. In the context of Australian medical school admissions, Griffin & Hu (2015) measured SEC using the Index of Relative Socio-Economic Advantage and Disadvantage (IRSAD), which is based on the 2011 Australian population census data and is derived from a composite of information on income, unemployment rates and home ownership. They found that final medical school GPA was *not* significantly correlated with ISRAD values. Simmenroth-Nayda & Görlich(2015) provide findings from the German context, and report that a group of applicants with parents who were medical doctors did not show any advantage in either test (multiple mini-interview (MMI) and interview), their individual preparation for the admission test, or in receiving or accepting a place at medical school. However, candidates with parents who were medical doctors had scored significantly lower in school exam grades, with a small effect size.

School Type

3.11. There are two key approaches to examine the effect of school type on attainment in higher education (see Ogg, Zimdars, & Heath, 2009 for a detailed discussion):

3.11.1. **School type effect:** school type effect examines the different degree performance of independent versus state school students who have obtained the same A-Level grades at the point of entry to HE. Generally, the literature shows that for a given set of A-Level grades, the degree performance of independent school students is lower than that of state school students.

3.11.2. **School performance effect:** school performance is defined as the average A-Level performance of all the students in a school. The school performance effect is a difference in degree performance between students who attended highly performing schools and students who attended schools performing less well, relative to the individual students' own A-Level results.

3.12. These two school effects will inevitably overlap to some extent; however, Smith & Naylor (2002) and HEFCE (2003) found a consistent school *type* effect, of a far larger magnitude than the school *performance* effect. Similar results confirming the school type effect have been found more recently in a cohort of English A-Level students who entered full-time degree courses in 2007-08 (HEFCE, 2014). The report found that state school students tend to do better in their degree studies than students from independent schools with the same prior educational attainment.

3.13. HEFCE (2014) also found evidence of school *performance* effect: when students with the same prior educational attainment are considered, those with A-Level grades that are better than the average for their school tend to attain more highly in HE than similar students with grades that are lower than the average for their school (school performance effect). An earlier HEFCE report (2003) found that where the school performance effect did exist, this was strongest in subjects allied to medicine and engineering (approximately three times stronger than the effects seen in the data overall, which covered many degree subjects).

3.14. As a case example of school effects, Garlick & Brown (2008) described the performance of widening access candidates on the Extended Medical Degree Programme (EMDP) at Kings College London. To be eligible for the EMDP, candidates must have attended one of the 100 or so state schools or colleges in one of the 15 most educationally deprived boroughs in inner London. Almost all the eligible schools performed below the national average (in some cases as much as 70% below) in examination results in 2005-6. When EMDP students first sit identical papers to their conventional peers (at the end of their second year on the programme), 10% (five students) routinely gain merit awards for coming in the top 15% of the whole year group. Garlick and Brown conclude that medical students can succeed with CCC grades at A-Level if their results are achieved at a low achieving school or college (school performance effect), although extra academic and pastoral support is needed to enable these students to reach their full potential. McManus et al (2013) found that students from higher attaining secondary schools performed less well in their first-year medical school exams than those from lower attaining secondary schools with similar A-level results.

3.15. It is usually assumed in the literature that the cause of a school effect is as follows (HEFCE, 2003): students who gain A-Levels at relatively disadvantaged schools do not show their full potential through their A-Level results. Once in HE, there is a relatively 'level playing field' and the students from state schools then achieve more than would be expected from their A-Level grades. Aside from school type and school performance effects, it is apparent that students from independent school students are generally found to enter higher education with better A-Level grades than those from state schools (HEFCE, 2014; Tiffin et al., 2014). There is some evidence that individuals from independent/grammar schools also perform better than others in all subtests of the UKCAT (James et al., 2010; Tiffin et al., 2014) (note again that this was before the inclusion of the SJT). Published evidence on the impact of school type on scores on **non-academic** admissions tests is limited, but Taylor et al. (2015) found no significant differences between applicants from selective and/or fee-paying schools, or neither selective nor fee-paying schools at the University of Birmingham's medical school, with offers given based on performance on an MMI.

Free School Meals (FSM)

3.16. Published research which addresses FSMs is sparse. The only research identified in this review was published by Crawford (2014), which found evidence of sizeable differences in university outcomes between pupils with and without FSMs, even amongst the selected group of university participants. Students who attended one of the 20% of secondary schools with the highest proportions of FSM-eligible pupils are, on average, 5.4% more likely to drop out, 11.0% less likely to complete their degree and 21.8% less likely to graduate with a first or a 2:1 than pupils who attended one of the 20% of secondary schools with the lowest proportions of FSM-eligible pupils.

Postcode and Participation of Local Areas (POLAR)

3.17. Researchers have acknowledged the likely confounds between postcode or statistics in local areas, and other measures of socioeconomic status such as parental education (e.g. Do et al., 2006) . Indeed, Steven et al.(2016) found that the majority of applicants to medical schools in all postcodes had parents in the highest SEC occupational group (NS-SEC1), implying that the use of postcode as a demographic marker for inequality may be misleading.

3.18. Nonetheless, some researchers have found significant results using postcode and local areas as predicator variables. Steven et al. (2016) found that applicants resident in the most deprived postcodes, with parents from lower SEC occupational groups (NS-SEC4/5) and attending non-selective state schools were less likely to obtain an accepted offer of a place at medical school (although this varied significantly by medical school). Similarly, the HFECE report in 2014 classified the postcodes students live in immediately prior to entry to HE using either the Income Deprivation Affecting Children Index (which measures in a local area the proportion of children under the age of 16 who live in low-income households) or POLAR (which measures in a local area the proportion of young people who go onto HE). Using either measure, those from the most disadvantaged areas were found to have consistently lower degree outcomes than those with the same prior educational attainment from other areas.

3.19. However, Taylor et al. (2015) used the POLAR3 quintile, which is based on the probability that a randomly selected young person from the applicants' area would participate in HE, based on the decisions of young people from that area between 2005/6 and 2010/11. No significant differences were found between groups of applicants to the University of Birmingham's medical school, with offers given based on performance on an MMI.

Ethnicity

3.20. In contrast to the demographic picture of university admissions decades ago, Crawford & Greaves (2015) reported that non-White students are increasingly likely to participate in HE, and are now significantly more likely to go to university than their White British counterparts. This promising trend is supported by Bodger et al. (2011) who found no significant relationship between ethnicity or nationality and medical school examinations in the first or final years of study at the University of Wales.

3.21. However, other research shows marked differences between ethnic groups in performance at the point of entry to HE, as well as for degree outcomes. For example, HEFCE (2014) report significant variation in degree outcome for students from different ethnicities. Students classifying themselves as White consistently achieved higher degree outcomes than students recording other ethnicities. In all, 72% of White students who entered higher education with BBB gained a first or 2:1 degree classification. In comparison, 56% of Asian students, and 53% of Black students gained these degree outcomes, entering with the same A-Level grades. Similarly, Haq et al. (2005) looked at performance data from two large medical schools and found that both men and women of Asian origin performed less well in undergraduate written and OSCE examinations than their White counterparts, although the effect size was not large. At the point of selection to medical school, (James et al., 2010) found that White students performed better than non-White students in all parameters on the UKCAT ($p<0.001$). Tiffin et al. (2014) also found that candidates reporting themselves as of White ethnicity, on average, achieved higher A-Level tariffs and UKCAT scores than those describing themselves as non-White. This effect was apparent even after controlling for the effect of other predictor variables.

Disability/Health Problems

3.22. Very little research has been published regarding the impact of disabilities and health problems on outcomes in higher education, potentially due to the highly complex and varied nature of these variables. However, one published study (Searcy, Dowd, Hughes, Baldwin, & Pigg, 2015), found that among applicants to American medical schools, those with Medical College Admissions Test (MCAT) scores obtained with extra test administration time due to disabilities, compared with standard administration time, had no significant difference in rate of medical school admission but had lower rates of passing the United States Medical Licensing Examination Step examinations and of medical school graduation within 4 to 8 years after matriculation. These performance differences persisted after controlling for MCAT scores and undergraduate grade point average.

4. Data Source

4.1. The data for this project includes information derived from HESA Student Record 2007/08 and 2008/09 collected by the Higher Education Statistics Agency Limited ("HESA") and provided to the GMC ("HESA Data"). The copyright for the data rests with HESA but the agency does not have any responsibility for the inferences and conclusions in this report. The data were de-identified and accessed through the Health Informatics Centre (HIC) Safe Haven ("Safe Haven User Guide", 2015).

4.2. All statistical analysis was performed using SPSS (version 23) and MLwiN (version 2.35).

5. Identifying cases for analysis

5.1. The base data set consisted of the two cohorts that started their medical education at a UK medical school in 2007 or 2008 (including both the standard entry to medicine together with Medicine with a Preliminary Year and Medicine with Gateway year). Only those who made sufficient progress and applied to foundation training for the first time in 2012, 2013 or 2014 were included. 13,182 cases were identified.

5.2. This data set includes UK educated students and those with other educational backgrounds. Because of the lack of widening access variables and the difficulty in equating educational qualifications for those educated outside the UK the study focused only on those educated in the UK. The study also focused on those who did not have a previous graduate qualification before starting medical school as there could be confounding factors in the qualifications and background for this group. Tables 5.1 shows the impact of these selection criteria on the sample. Overall 24% of the two cohorts are excluded.

Table 5.1 Data counts by HESA year of commencement at medical school and place of education

	Place of education				
HESA Commence Year	Secondary and Undergraduate Medical Education in the UK		Secondary and/ or Undergraduate Medical Education outside the UK	No information on place of education	Total
	Non-Graduate Entry	Graduate Entry			
2007	5,089	488	508	234	6,319
2008	4,947	1,039[2]	577	300	6,863
Total	10,036	1,527	1,085	534	13,182

5.3. Previous research examining the progress of undergraduate students at university (Chowdry et al., 2013; HEFCE, 2014) as well as progress of students at primary and secondary schools (e.g. DFE, 2016, Thomas, 2010, Munoz-Chereau & Thomas, 2016) typically uses a sample based on the cohort completing a stage of education and obtaining a particular qualification (e.g. outcome measure such as degree outcome, A-level, GCSE). Essentially this is because it is important to better align equivalent student experiences in different institutions by comparing the relative effectiveness of teaching and learning practices for the students obtaining an outcome qualification at the same time point.

5.4. However, this approach was not used in the current analysis due to limitations of the available dataset. Nevertheless, in due course, once the relevant data becomes available it is anticipated that for equivalent future analyses the key sample will be defined using the year of graduation from medical school cohort rather than the year of entry, as the next stage of UKMED will contain data on all those starting medical school between 2007 and 2014.

[2] The change in graduate figures between 2007 and 2008 reflects the fact that two further schools initiated a graduate entry track.

Predictor Variables

5.5. **HESA Tariff** - The HESA tariff assigns points to educational qualifications allowing comparison between different qualifications and educational routes. The methodology for the tariff has recently been amended but the regime used in 2007 and 2008 was the same. A grade B A-level is awarded 100 points, thus three B grade A-levels gives a total score of 300. The selectivity of entrance to medical school makes it unlikely that someone with this level of achievement would receive an offer, even in a widening access context. 659 records had scores of 300[3] or below and these cases were dropped from the study leaving 9,377 cases with valid values. Table 5.2 summarises HESA Tariff scores for the full sample and for the selected cases. A t-test showed no significant difference between the 2007 and 2008 cohorts for the restricted score (t=0.53, df=9375, ns).

Table 5.2 Summary of HESA tariff scores

	N	Minimum	Maximum	Mean	Standard Deviation
All cases	13,182	0	1036	395	207.2
All Selected Cases	10,036	0	1036	466	136.1
All cases Tariff>300 ,	10,369	310	1036	494	84.2
Tariff>300 , Selected	9,377	310	1036	495	82.8

5.6. The HESA tariff scores have a wide distribution reflecting different educational experiences. While typically students take 3 A-levels resulting in a maximum possible tariff of 360, others take more with 5 or 6 not uncommon in the data. Some students may have sat additional A-level qualifications in an extra year – perhaps to improve grades or to add to their selection of subjects to better match course requirements. Records for candidates with very high HESA tariff scores were inspected but there was no evidence of error. Sensitivity analyses were conducted to consider capping high tariff scores at 800, but there was no impact on the results, so the original scores were used. (See Appendix B).

5.7. The total HESA tariff is not the only approach for creating a single score from secondary education qualifications, however it is one which is well researched and commonly used. Other approaches for scoring secondary qualifications do exist. Many universities use the score from the top 3 A-levels to create a tariff and sometimes scores from General Studies are discounted. McManus et al (2013) in their study of the validity of the UKCAT in 12 universities (UKCAT12) use an approach like this. Sensitivity analyses were conducted to compare the UKCAT12 approach with the use of the full HESA score. Details of these analyses are provided in Appendix B.

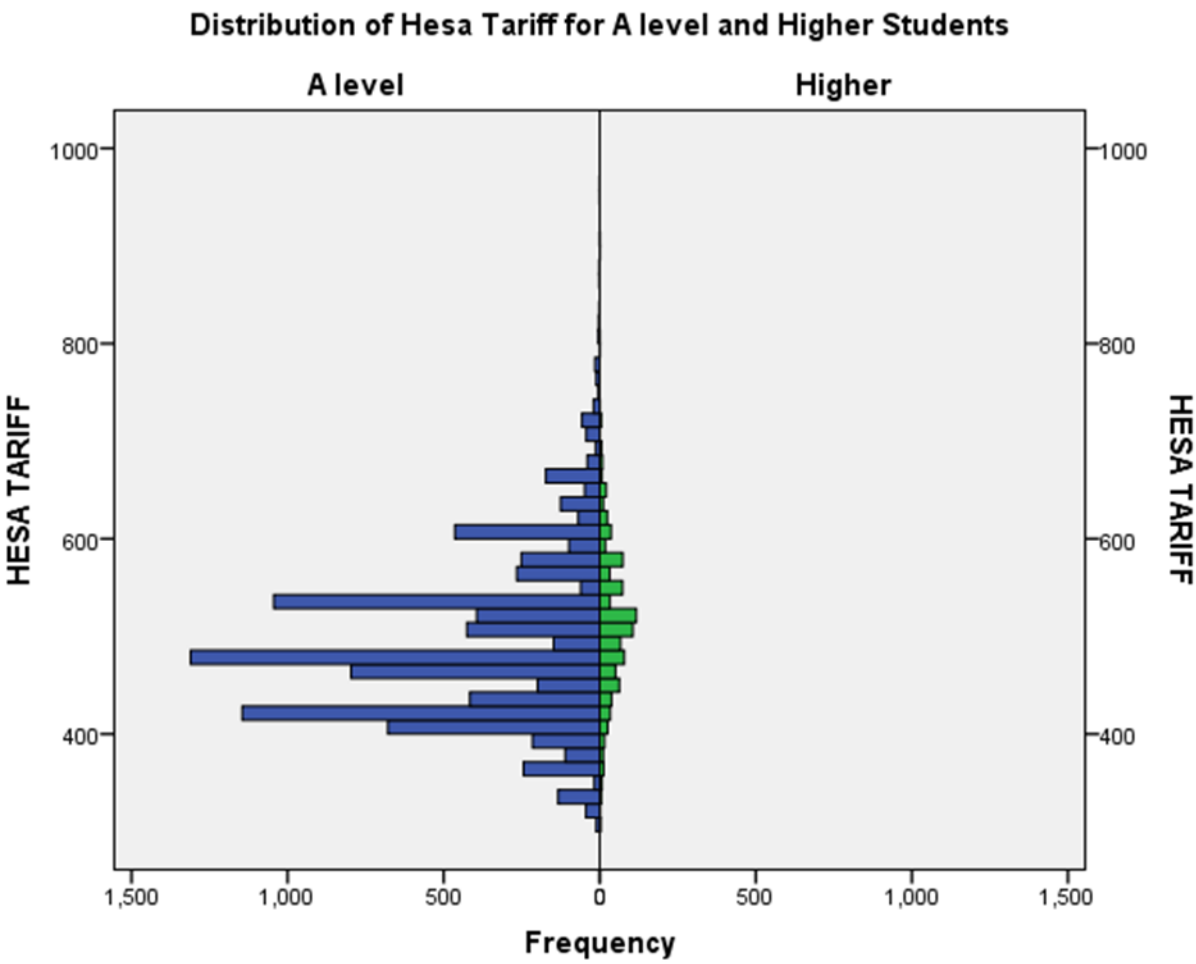
5.8. In this case there was some evidence for improved prediction of outcomes using the UKCAT12 approach, however this required the removal of all cases with qualifications other than A-levels and Highers from the analysis. This might have disproportionately removed those from widening participation groups who had qualified through other routes. In addition there was some evidence that the approach was not as successful as

[3] The assumption was that records with fewer than 300 HESA points were incomplete and therefore the tariff was incorrect and did not accurately reflect the academic attainment of the candidate.

the HESA Tariff in equating between the different qualification regimes of Scotland and the rest of the UK. For these reasons the HESA tariff scores was used.

5.9. Figure 5.1 shows the distribution of HESA Tariff scores for A-level and Higher students separately. There is somewhat less variance on the Higher distribution although the sample is smaller and therefore fewer extreme scores would be expected. The centres of the two samples are well aligned. There are clusters of A-level scores which represent different numbers of A-level passes.

Figure 5.1 Population of pyramid of HESA tariff scores broken down by qualification type



5.10. **UKCAT** - The UKCAT consisted of four cognitive ability tests at the time these cohorts sat the tests. (The UKCAT SJT was introduced in 2013). The four tests are:

- Abstract Reasoning
- Decision Analysis
- Quantitative Reasoning
- Verbal Reasoning

5.11. Scores for each test are reported on a standardised scale running from 300 to 900 with a mean of 600 and a standard deviation of 100. Test forms are equated from year to year so that scores from different years of taking the tests can be considered equivalent. Applicants can take the test again in following years if their application was unsuccessful in one year and a number of candidates had done so. 704 candidates had taken the test twice and 4 individuals had done so three times. For the purpose of the analysis the last recorded UKCAT score was used. 702 of the 10,036 cases had no UKCAT scores. Table 5.3 summarises the UKCAT scores for this group and shows the inter-correlation of the scores. The mean scores are all above 600 and the standard deviations less than 100. This reflects the restriction of range that would be expected due to selection on the test scores. The correlations between the different subtests, ranging from 0.20 to 0.31, are moderate. This supports the use of the independent test scores rather than using a combined score.

5.12. Independent sample T-tests showed no significant difference between the 2007 and 2008 cohorts for the Decision Analysis and Verbal Reasoning Tests (t=0.96 and 1.76). There were small but significant differences on the other two tests between the cohorts with the 2007 cohort performing better on the Abstract Reasoning measure (t=6.0, df=9201, p<0.01) and the 2008 cohort performing better on the Quantitative reasoning test (t=15.4, df=9332, p<0.01). The largest difference was for the Quantitative reasoning test with a score difference of just under a quarter of a standard deviation.

Table 5.3. UKCAT scores and intercorrelation (n=9334)

Correlations					
Test	Mean	SD	Decision Analysis	Quantitative Reasoning	Verbal Reasoning
Abstract Reasoning	623.7	78.7	0.29	0.25	0.20
Decision Analysis	627.2	92.9		0.27	0.26
Quantitative Reasoning	639.4	70.5			0.31
Verbal Reasoning	623.1	78.7			

5.13. A sensitivity analysis was conducted to examine the utility of using a combined UKCAT score as a predictor variable, rather than the individual UKCAT sub-test scores. These analyses indicated that using the individual UKCAT sub-test scores had significantly higher predictive power compared to a combined score. Checks for collinearity showed no issue with using the four separate scores so this was the preferred approach. See Appendix C further details of the sensitivity analysis.

Outcome Variables

5.14. Three outcome variables were considered. ARCP ratings were available for the trainees' performance during their F1 and F2 clinical training years. In addition, trainees completed applications on the Foundation Programme Application System (FPAS) assessments which provided the EPM decile and SJT scores. These scores are described below:

5.15. **ARCP scores** – These scores reflect performance in clinical training and as such could be considered to represent the culmination of medical school training where the primary focus is to produce competent professionals. Differences in ARCP scores should therefore reflect how well students have progressed during their studies and at an institutional level should reflect the quality of training students have received.

5.16. ARCP scores were ordered according to the approach suggested by Tiffin et al (2014). There was relatively little variance in the ARCP scores with the majority of individuals receiving the maximum score of "satisfactory" progression. Table 5.4 shows the score distributions for year 1 and year 2 ARCP ratings. It can be seen that most frequent non-satisfactory category is "Insufficient Evidence". This may be an indicator of poorer performance but may also result from poor record keeping on the part of either the trainees or their supervisors.

Table 5.4 ARCP scores

ACRCP Outcome ("Tiffin score")	Year 1		Year 2	
Commence Year	2007	2008	2007	2008
Satisfactory	4563	4421	4179	1988
Insufficient Evidence	488	482	520	270
Targeted training required without extension of training time			1	0
Extended training time required or left programme	25	33	34	17
Missing	24		3018	

5.17. **EPM Scores** – The EPM ("Educational Performance Measure", 2016) score is based on assessments during medical school training. It represents the academic achievement of students on their course in contrast to the ARCP scores which focus on clinical practice. Typically in value-add studies, academic attainment variables are used as output measures. For example in studying value-add in Secondary schools the outcome measure might be GCSE or A-level scores. The available predictor measures are well matched for predicating academic performance since they are a mix of cognitive measures and prior academic performance.

5.18. The EPM score is expressed as a Decile score comparing students within school. Each medical school measures the performance of their students using their own approach. The results for each medical school are split into ten equal groups ranked by performance. The performance is compared within year cohort only. The lowest Decile is assigned a score of 34 and the highest Decile a score of 43. The schools use different methods and assessments for calculating these scores which are made available to their own students and may be published on the schools' websites. Because of the way the EPM score is calculated it cannot be used to compare across schools since between school variance has been controlled during the calculation of scores – the ranking is within school and year cohort, not across the UK.

5.19. The EPM score is used as part of the FPAS application process and candidates can have multiple applications. Where a candidate has multiple EPM scores across the years the final score was used. Therefore, although the sampling focused on 2013 and 2014 applications for placements, a few candidates re-applied in 2015.

5.20. Table 5.5 summarises EPM scores for the two cohorts across different application years. The 2007 cohort would have applied mostly in 2012 and 2013. Those assessed in 2014 and 2015 have lower EPM scores suggesting these are individuals applying again after having been unsuccessful in previous year. A similar pattern is seen for the 2008 cohort with lower scores in 2015 which are all repeat applications.

Table 5.5 EPM scores

FPAS Year	Commence year = 2007		Commence year = 2008	
	Median	N	Median	N
2013	39.0	3015	38.4	3123
2014	35.9	297	39.0	2953
2015	34.9	43	35.0	87
Total	38.7	3355	38.7	6163

5.21. **FY1 SJT Scores -** The FY1 SJT is a machine marked test which was first used as part of the FPAS process in 2013. The SJT is designed to measure the professional attributes associated with being an F1 doctor such as Commitment to Professionalism, Coping with Pressure, Effective Communication, Patient Focus and Working Effectively as Part of a Team. These should have been developed during education and training so performance on this measure reflects an aspect of the quality of training. Further information about the test is provided in Appendix F.

5.22. The FY1 SJT is designed to be a less cognitively loaded measure than the prior attainment measures, however it does capture learning from the course. It is typically positively correlated with cognitive measures but not strongly so. Verbal loadings are highest as the content is provided as written descriptions of situations that arise in clinical practice. Table 5.8 shows the SJT correlated 0.24 with the Verbal UKCAT Subtest and between 0.10 and 0.15 with the other subtests.

5.22.1. Scores for multiple forms are equated within year but not across years, therefore equated scores are not fully equivalent across application years. As for the EPM, where a candidate has multiple FY1 SJT scores across the years the final score was used. Table 5.6 shows the FY1 SJT scores for the studied cohorts. The sample sizes are smaller than for the EPM because of missing values within the data set. Table 5.7 summarises the SJT score results for the whole cohort that sat the test in each year taken from the test report published by The Foundation Programme. Given the large differences between mean scores across years, SJT scores were converted to z-scores based on the performance of the whole cohort for the year the test was taken.

Table 5.6 Within year equated FY1 SJT scores for study sample

FPAS Year	Commence year = 2007			Commence year = 2008		
	Mean	SD	N	Mean	SD	N
2013	866	24	2,677	862	25	2,318
2014	835	28	223	842	23	2,592
2015	873	26	2			

Table 5.7 FY1 SJT scores for full cohort taking the test

FPAS Year	Mean	SD	N
2013	858.7	31.5	8,163
2014	836.4	27.9	7,957
2015	887.6	29.8	8,088

Relationship between Predictors and Outcome variables

5.23. The analysis will look at how the relationship between predictor and outcome variables varies for different medical schools or when various demographic variables are controlled. For this to be effective a substantial base relationship between predictor and outcome variables is desirable. Table 5.8 shows the raw correlations between the different variables.

5.24. The ARCP variables show almost no relationship with the predictor tests. Negative correlations would be expected given the scoring convention for the ARCP with 1 being the best score and 4 the lowest.

5.25. The EPM and the FY1 SJT both show statistically significant relationships with the predictors with the SJT showing the strongest relationships. This makes the FY1 SJT score the preferred outcome variable both because it has stronger relationships with the predictors than the EPM and because the EPM does not allow comparison between medical schools because of the way it is calculated removing between school variance[4]. **On this basis, despite the limitations described above, the FY1 SJT was chosen as the key outcome measure in Multilevel Modelling (MLM) analyses.**

Table 5.8 Correlations between predictor and outcome variables

Outcome Variables Predictors	**ARCP Year 1~**	**ARCP Year 2~**	**FPAS EPM#**	**FY1 SJT~**
HESA Tariff	-.01 n=10,012	-.02 n=7,017	0.13** n=7,371	0.10** n=7,371
UKCAT Abstract Reasoning	-0.02* n=9,312	-.02 n=6,425	0.11** n=7,365	0.13** n=7,364
UKCAT Decision Analysis	-.02 n=9,312	.00 n=6,425	0.11** n=7,365	0.15** n=7,364
UKCAT Quantitative Reasoning	-.01 n=9,312	.00 n=6,425	0.06** n=7,365	0.13** n=7,364
UKCAT Verbal Reasoning	-.00 n=9,312	.00 n=6,425	0.20** n=7,365	0.24** n=7,364

~ Pearson correlations; # Spearman Correlation * p<0.05 two tailed; ** p<0.01 two tailed

[4] Moreover, any medical school differences identified using EPM outcome may to some extent reflect differences in HEI assessment standards rather than the reality of student progress. Typically in value added methodology for schools both the prior attainment and outcome assessment should be equivalent across institutions. Externally assessed examinations (eg GCSE, A-level) provide suitable measures.

5.26. Both the HESA Tariff and the UKCAT tests predict FPAS FY1 SJT scores. The strongest predictor is the UKCAT verbal reasoning measure ($r=0.24$, $p<0.01$).

Selected Sample and Background Variables

5.27. There are 6,978 cases with values for the five predictor variables and the FY1 SJT that meet the additional criteria described above. Table 5.9 shows the summary values for the variables for the selected sample with all the chosen background variables (demographic sample) as well as the variables required for the prediction model (demographic sample). It also shows the results for the largest sample with full data for the prediction model alone (prediction sample) and the full sample (based on cohort starting medical school see table 4.1). UKCAT scores and the SJT are reported as z-scores. For the UKCAT the standardisation is based on the full sample. For the SJT the results for the full cohort that took the test originally are used to convert the scores to z-scores. For consistency, in subsequent MLM analyses the HESA tariff is also standardised to a z-score metric based on the full sample.

Table 5.9 Summary variable values

Variable	Demographic Sample All demographics n=4691		Prediction Sample Main model variables n=6,978		Full Sample All available for variable		
	Mean	SD	Mean	SD	Mean	SD	n
HESA Tariff	496.6	82.6	497.4	84.2	493.9	84.2	10,369
Z HESA Tariff	0.03	0.98	0.04	1.00	0.00	1.00	10,369*
UKCAT Abstract Reasoning	622.3	78.1	624.6	78.6	621.5	80.1	11,972
Z UKCAT Abstract Reasoning	0.01	0.98	0.04	0.98	-0.00	1.00	11,972
UKCAT Decision Analysis	632.4	91.7	631.4	92.0	627.0	93.8	11,972
Z UKCAT Decision Analysis	0.06	0.98	0.05	0.98	0.00	1.00	11,972
UKCAT Quantitative Reasoning	646.1	69.1	645.1	69.8	639.1	71.8	11,972
Z UKCAT Quantitative Reasoning	0.10	0.96	0.08	0.97	0.00	1.00	11,972
UKCAT Verbal Reasoning	625.5	77.9	625.1	78.5	621.8	81.1	11,972
Z_UKCAT Verbal Reasoning	0.05	.96	0.04	0.97	0.00	1.00	11,972
z_ FY1 SJT	0.19	0.80	0.18	0.80	0.13	0.83	9,511
Age at Start of Medical School	18.3	0.70	18.3	0.73	19.4	2.9	13,182

*Standardisation is based on all records with tariff score above 300, whether they met other inclusions criteria or not.

5.28. The age at start of medical school shows that the selected samples are a little younger than the full sample. This is likely to be because overseas and graduate applicants are older than UK school leavers. They are excluded from this analysis but included in the full sample.

5.29. **Demographic and Socio-Economic Variables -** Table 5.10 shows the breakdown of the selected sample and the full sample by different demographic and socio-economic variables. There are only minor differences between the Demographic and Prediction Samples and the full sample on the background variables.

Table 5.10 Demographic variables

Variable	Demographic Sample All demographics n=4691		Prediction Sample Main model variables			Full Sample All available for variable		
	Count	Percentage of valid answers	Count	Percentage of valid answers	N valid responses	Count	Percentage	N valid responses
Female	2,658	56.7%	3883	55.6%	6,978	7528	57.1%	13,182
BME (any non-white group)	1,294	27.6%	2,174	31.2%	6,963	4167	31.7%	13,139
Reported a disability	45	1.0%	63	0.9%	6,884	137	1.1%	12,776
At least one parent with a degree	3,746	79.9%	4,943	73%	6,782	7859	70%	11,260
Managerial and Professional Occupations	3,550	75.7%	5,095	75.1%	6,786	4761	46%	10,447
Low participation neighbourhood (Polar 2)	169	3.6%	257	3.8%	6,853	513	4.3%	11,925
Attended Private School	1,420	30.3%	2,202	32%	6,851	3339	29%	11,658
Received Free School Meals and/or Income Support	724	15.4%	1,020	16.4%	6,231	1749	16.9%	10,347

5.30. **Breakdown by medical school -** Most candidates attended the same medical school for the whole of their training. A few individuals changed school during their training. Two schools that offer initial medical education transfer their whole cohort to another school to complete their education. For the purposes of this analysis an individual was always assigned the final school attended where there was more than one. Clinical placements which are most relevant to developing the competencies measured by the FY1 SJT are predominantly undertaken in the later years of the course. Table 5.11 shows the breakdown by Medical school. Most schools represent a similar proportion of the full sample as of the selected sample. Where schools have a higher percentage in the demographic and prediction samples than in the full sample it is likely that they have fewer graduate entrants and students from overseas and conversely where the selected sample constitutes a smaller proportion of the school students the school may take more students from overseas or be one of the schools that did not require applicants to complete the UKCAT tests in 2007 and 2008.

Table 5.11 Breakdown by medical school

	Demographic Sample All demographics n=4691		Prediction Sample Main model variables n=6978		Full Sample n=13182	
Medical School	Count	Percent of sample	Count	Percent of sample	Count	Percent of sample
Aberdeen	139	3.0%	159	2.3%	337	2.6%
Barts	157	3.3%	286	4.1%	573	4.3%
Birmingham	310	6.6%	383	5.5%	720	5.5%
Brighton and Sussex	19	0.4%	123	1.8%	246	1.9%
Bristol	199	4.2%	259	3.7%	432	3.3%
Cambridge	120	2.6%	183	2.6%	267	2.0%
Cardiff	159	3.4%	317	4.5%	638	4.8%
Dundee	145	3.1%	175	2.5%	288	2.2%
Edinburgh	233	5.0%	286	4.1%	478	3.6%
Glasgow	205	4.4%	251	3.6%	480	3.6%
Hull/York	64	1.4%	105	1.5%	260	2.0%
Imperial	222	4.7%	438	6.3%	630	4.8%
Keele	78	1.7%	96	1.4%	237	1.8%
King’s	267	5.7%	376	5.4%	676	5.1%
Lancaster	31	0.7%	41	0.6%	85	0.6%
Leeds	125	2.7%	292	4.2%	456	3.5%
Leicester	103	2.2%	177	2.5%	391	3.0%
Liverpool	219	4.7%	278	4.0%	547	4.1%
Manchester	352	7.5%	433	6.2%	784	5.9%
Newcastle	242	5.2%	293	4.2%	599	4.5%
Norwich	72	1.5%	103	1.5%	297	2.3%
Nottingham	155	3.3%	190	2.7%	548	4.2%
Oxford	97	2.1%	207	3.0%	277	2.1%
Peninsula	136	2.9%	182	2.6%	361	2.7%
Queen’s	148	3.2%	248	3.6%	475	3.6%
Sheffield	183	3.9%	228	3.3%	412	3.1%
Southampton	142	3.0%	192	2.8%	410	3.1%
St George’s	151	3.2%	221	3.2%	436	3.3%
UCL	218	4.6%	456	6.5%	669	5.1%
Warwick	0		0		173	1.3%

6. Univariate Analysis of Demographic Variables

6.1. This analysis considers differences in the outcome variable that are related to the different demographic factors. Because the data are static it is not possible to determine that the demographic variables are the cause of any differences found or are confounded with other causal variables. For example, a finding that female students performed better might be caused by higher levels of aptitude among women applicants, but it may be that better schools were more likely to accept or attract female applicants. However, these findings indicate which demographic variables may be of interest for further analysis.

6.2. Where the demographic variable takes two values, a two-tailed independent groups t-test was performed. Where there were more than two groups, a one-way ANOVA was used to test for significant differences. A correlation is reported where the demographic variable is treated as interval. The analysis was repeated for the prediction sample (upper line in table 6.1) and the demographic sample (lower line in table 6.1).

6.3. Table 6.1 shows the results. There were significant findings for all variables except disability. It was decided to include all variables in the multilevel analysis since this approach is likely to be more sensitive to effects than the univariate approach.

Table 6.1 Differences in HESA, UKCAT and FY1 SJT scores by group

Demographic Variable	HESA Test Statistic (df) p value	UKCAT (total) Test Statistic (df) p value	FY1 SJT Test Statistic (df) p value	Comments on effect
Age	r=-0.16 (6977) .00	r=-.13 (6977) .00	r=-.08 (6977) .00	Younger individuals perform better on all measures.
	r=-0.17 (4690) .00	r=-.13 (4,690) .00	r=-0.07 (4690) .00	
Medical School	F=33.2 (28,6949) .00	F=44.5 (28,6949) .00	F=11.2 (28,6949) .00	Significant differences between schools on all variables.
	F=20.1 (28,4662) .00	F=30.5 (28,4662) .00	F=8.2 (28,4662) .00	
Gender	t=3.35 (6466) .00	t=4.05 (6976) .00	t=10.7 (6976) .00	Men have higher HESA and UKCAT scores but women have higher scores on the FY1 SJT by a fifth of an SD.
	t=3.2 (4689) .00	t=3.8 (4689) .00	t=8.7 (4689) .00	
BME Status	t= 0.5 (6961) ns	t=8.2 (6961) .00	t=13.6 (6961) .00	The BME group scored about one quarter of a standard deviation lower on both the UKCAT and FY1 SJT.
	t=0.5 (4689) NS	t=6.2 (4689) .00	t=10.2 (4689) .00	
Reported Disability	t=1.5 (6882) ns	t=1.3 (6882) ns	t=1.0 (6882) ns	No significant differences.
	t=1.1 (4689) ns	t=0.56 (4689) ns	t=0.44 (4689) ns	
At least one parent with a degree	t=7.32 (3488) .00	t=10.2 (6780) .00	t=3.3 (6780)) .00	Those whose parents have degrees score higher on average on all indicators. The largest difference is about a quarter of an SD on the UKCAT total score.
	t=4.1 (4689) .00	t=5.3 (4689) .00	t=3.3 (4689) .00	

Demographic Variable	HESA Test Statistic (df) p value	UKCAT (total) Test Statistic (df) p value	FY1 SJT Test Statistic (df) p value	Comments on effect
Parents' Profession (SEC_Combined)	F=6.3 (4,6781) .00	F=17.4 (4,6781) .00	F=5.6 (4,6781) .00	Individuals whose parents were not in managerial and professional occupations had higher scores on UKCAT and FY1 SJT. The trend was similar with the HESA tariff but lower supervisory and technical occupations had the highest HESA scores on average.
	F=4.7 (4,4686) .00	F=10.3 (4,4686) .00	F=3.6 (4,4686) .00	
Low participation neighbourhood (Polar 2)	t=0.3 (6851) ns	t=3.0 (6851) .00	t=2.1 (6851) ns	The significant differences for UKCAT and FY1 SJT show lower scores in Low participation neighbourhoods but are not consistent across samples.
	t=0.79 (4689) ns	t=1.6 (4689) ns	t=2.1 (4689) .04	
Attended Private School	t=4.3 (6849) .00	t=3.8 (6849) .00	t=0.3 (6849) ns	Those from state funded schools have higher HESA tariffs but lower UKCAT scores.
	t=3.6 (4689) .00	t=3.2 (4689) .00	t=0.0 (4689) ns	
Received Free School Meals and/or Income Support	t=2.8 (6229) .01	t=7.9 (6229) .00	t=6.1 (6229) .00	Those with families receiving income support performed less well on all but one comparison by up to a quarter of an SD.
	t=1.7 (4689) ns	t=6.4 (4689) .00	t=4.5 (4689) .00	

7. Modelling Medical School Differences

7.1. In the modelling process the sequential impact of various predictor variables in explaining variance in the outcome variable is tested. Multi-level modelling (MLM) shows the impact of different variables across the whole sample and in differentiating schools.

7.2. Independent student outcome variable for all the models is the FY1 SJT score expressed as a z-score, standardised within year based on the full cohort for that year. Two clustering variables are used in all MLM analyses to partition the variation in SJT outcomes that can be attributed to differences between students or alternatively differences between schools.

7.3. Four models are presented with the aim to identify what improvement in prediction is attained with increasing complexity of the models.

- **Null model – including no explanatory variables**
- **Prior attainment only model**
- **Prior attainment plus student background characteristics model**
- **Prior attainment plus significant student background characteristics model**

7.4. The initial analyses included the starting year cohort (2007 or 2008) and the year of taking the FY1 SJT test as control variables. These variables consistently showed zero effects and so were dropped from the models presented here. Where candidates sat the SJT test more than once, the last sitting was used as the operational variable.

Null model - including no explanatory variables

7.5. The first model is **without explanatory variables that provides a baseline for comparison**. Only the Medical school is included as a cluster variable. The model was estimated for the maximal sample – all individuals with FY1 SJT scores, for the predictor sample with all variables in the base model and for smaller demographic sample with all the background variables as well. The results are similar in all three cases. This suggests that the selected sample with full data is not substantially different from the records that had to be excluded due to missing data. Another illustration of this is shown in Appendix D which describes the results of using imputation to replace missing values.

7.6. The baseline models (table 7.1) show a significant reduction in the log likelihood values compared to a model with no variables showing that the introduction of the medical school has significant explanatory value for the SJT scores. This means there are some differences between schools in the FY1 SJT scores of their students. However, the value of the Variance Partition Coefficient (also referred to as the Intra-Class Correlation or ICC) suggests medical school attended accounts for 4% of the variance in the FY1 SJT scores. By way of comparison, studies exploring the impact of school on performance in secondary education typically find it accounts for 7-10% of the variance (Goldstein & Thomas, 1996; O'Donoghue, Thomas, Goldstein, & Knight, 1997; Thomas et al, 2013). However it is important to note that the outcome variables for secondary education are very different from the SJT variable used here so these findings are not directly comparable.

Table 7.1 Baseline model for three samples

Baseline Model	Demographic Sample n=4,691	Prediction Sample n=6,978	Maximal data set with FY1 SJT scores n=9,511
-2*log likelihood	11118.1	16474.1	23236.9
Reduction in -2*log likelihood	139.1	412.5	264.2
Intercept (se)	0.190 (0.034)	0.165 (0.032)	0.109 (0.032)
Variance (SE) Med School	0.028 (0.009)	0.026 (0.008)	0.027 (0.008)
Residual Variance (SE) student	0.619 (0.013)	0.615 (0.010)	0.669 (0.010)
Total Variance	0.647	0.641	0.696
Variance Partition Coefficient (ICC)	0.043	0.041	0.039

Prior attainment only model

7.7. This model **introduces the student predictor prior attainment variables of the HESA tariff and the four UKCAT scores**. Prior attainment variables are the most important variables used to calculate institutional value-added scores. By including these variables there is a further significant reduction in the error variance. The Verbal Reasoning test is the best predictor of the FY1 SJT scores. The HESA Tariff parameter is approximately equal to its standard error meaning that there is no significant contribution of the HESA Tariff to the prediction of SJT scores. In this model, controlling for students' prior attainment score on entry to medical school explains over half (54-58%) of the observed variance between medical schools. The overall goodness of fit (indicated by the total variance explained) is not large (6-7%) suggesting other unmeasured factors may also be important in predicting individual student SJT outcomes.

7.8. Moreover, having controlled for students' prior attainment, the ICC value indicates that the percentage of remaining variance in students SJT scores attributable to differences between medical schools is reduced and very small (2%). This finding is reflected in Figure 7.1 a caterpillar plot of value-added scores. It shows the average residual performance differences of students in each medical school after controlling for prior attainment (each medical school is denoted by a triangle) and the confidence intervals for the value (denoted by a line through the triangle). Where the confidence interval does not overlap the zero line the performance of students from the school is significantly different from the sample as a whole, even when prior attainment is controlled. The plot shows substantial overlap between medical schools in the confidence intervals for value-added scores (i.e. residuals from this model). However, around a third of schools have values which are significantly lower or higher than the overall sample values. **Higher scores mean that students from these schools are achieving somewhat better than students of similar initial ability attending other schools. Lower scores mean that students from these schools are achieving somewhat lower than students of similar initial ability attending other schools.**

7.9. One possible explanation of the observed low general model fit is that the relationship between prior attainment and FY1 SJT scores is not linear whereas all the models tested so far assume linear relationships. To examine whether non-linear models would show better fit, squared and cubed terms of prior attainment measures were added to the models. These terms provide tests of fit for non-linear trends. However, the addition of these variables to the prior attainment only model did not result in any significant improvement in the variance explained and so they have not been included here. This suggests that a linear model is the most appropriate for modelling the relationship. Details of the analysis are provided in Appendix E.

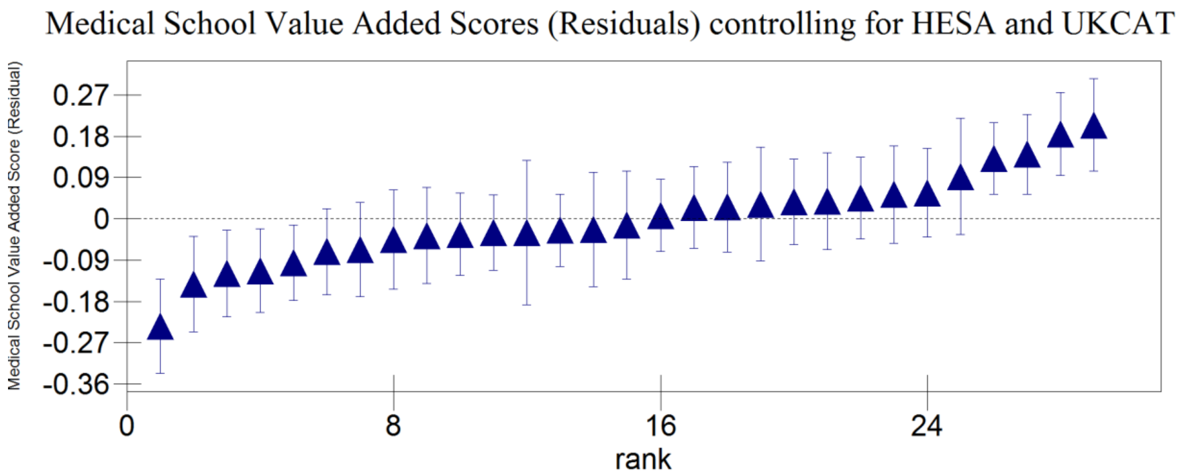
Table 7.2 Prior attainment model for 2 samples

Model	Demographic sample n=4,691	Prediction sample n=6,978
-2*log likelihood	10904.8	16101.1
Reduction in -2*log likelihood	213.3	373
Intercept (se)	0.182 (0.024)	0.163 (0.022)
Parameter HESA Tariff	0.010 (0.012)	0.015 (0.010)
UKCAT Abstract Reasoning	0.042 (0.012)	0.044 (0.010)
UKCAT Decision Analysis	0.037 (0.013)	0.048 (0.010)
UKCAT Quantitative Reasoning	0.030 (0.013)	0.023 (0.010)
UKCAT Verbal Reasoning	0.141 (0.013)	0.149 (0.010)
Variance (SE) Med School	0.013 (0.004)	0.011 (0.004)
Residual Variance (SE) student	0.593 (0.012)	0.584 (0.010)
Total Variance	0.606	0.595
Med School % Variance explained	54%	58%
Total % Variance explained	6%	7%
Variance Partition Coefficient (ICC)	0.021	0.018

The specific results for medical schools need to be treated with caution because this is an exploratory study and in particular, the weakness of the goodness of fit of the model, (shown by the low Variance Partition Coefficient) means that even though some 50% of school variance is explained by the prior attainment data, this is half of a relatively small amount. However, the plot does show how it is possible to compare Medical schools in terms of value-added since we can identify those with a value-added score that is statistically significantly different than would be expected from the cohort as a whole.

7.10. While we go on to examine the average impact of background variables on student outcomes in the next section, it was not appropriate to control for background factors in estimating school performance, except as they relate to prior attainment. To do so would build in expectations about demographic and SES group differences in estimating differences in school performance which was not considered appropriate for the purpose of this study (agreed by project advisory group June 2016).

Figure 7.1 Residuals by medical school for prior attainment model



Prior attainment plus student background characteristics model

7.11. In the final stage of the analysis, **student background variables and indicators of widening access were added to the model** to examine the typical impact of these variables on student outcomes, having controlled for prior attainment. In table 7.3, a positive sign to the reported parameter means that the comparison group tend to have higher scores on the SJT compared to the reference group and a negative sign means the reference group tend to have higher scores. This information is particularly important to guide support measures for disadvantaged groups at medical school. There is a small but statistically significant improvement in model fit (indicated by reduction in the -2ll ratio) compared to the previous analysis without the demographic variables, indicating a small impact of background variables. The loadings of the main predictors remain similar.

7.12. Of the background variables the strongest loadings are for gender and BME status and these two variables account for the majority of the improvement in prediction. The age and free school meals/income support variable showed small effects with older students and those who received free school meals or income support tending to have marginally lower performance. The loadings for the other background variables did not reach statistical significance.

7.13. Table 7.3 compares the results in the demographic sample for all the demographic variables and the restricted set. The final column shows the results for the model with restricted demographic variables on the largest sample possible for that set of variables. The first two models can be compared to the models in Table 7.2 with the same sample. The third model is not directly comparable with previous models, but uses the maximal data set possible which enhances the statistical power of the analysis.

Table 7.3 Prior attainment model with background variables, three treatments

Model	Demographic Sample n=4,691	Demographic Sample n=4,691	Prediction sample including background variables with significant effects. n=6,224
-2*log likelihood	10715.3	10718.3	14142.3
Reduction in -2*log likelihood	189.5	186.4	n/a
Intercept (se)	0.791 (0.303)	0.805 (0.303)	0.725 (0.251)
Parameter HESA Tariff	0.009 (0.012)	0.011 (0.012)	0.021 (0.010)
UKCAT Abstract Reasoning	0.036 (0.012)	0.035 (0.012)	0.029 (0.011)
UKCAT Decision Analysis	0.026 (0.012)	0.026 (0.012)	0.033 (0.011)
UKCAT Quantitative Reasoning	0.046 (0.013)	0.046 (0.013)	0.045 (0.011)
UKCAT Verbal Reasoning	0.121 (0.013)	0.121 (0.013)	0.126 (0.011)
Background Variables			
Age	-0.037 (0.016)	-0.037 (0.016)	-0.033 (0.014)
Gender (reference =male)	0.220 (0.023)	0.221 (0.23)	0.228 (0.20)
BME status (reference=white)	-0.227 (0.027)	-0.228 (0.027)	-0.238 (0.023)
Disabled Status (reference=not disabled)	0.028 (0.113)		
Free school meals / income support (reference=neither)	-0.072 (0.032)	-0.076 (0.031)	-0.091 (0.026)
Parents education (reference=degree level)	-0.022 (0.031) ns		
Parents occupation (reference=Manager)	-0.006 (0.028) ns		
School type	0.036 (0.025)		
Low participation neighbourhood (reference= not low participation)	-0.032 (0.060) ns		
Variance (SE) Med School	0.015 (0.005)	0.015 (0.005)	0.013 (0.004)
Residual Variance (SE) student	0.569 (0.012)	0.570 (0.012)	0.564 (0.010)
Total Variance	0.584	0.585	0.577
Med School % Variance explained	46%	46%	50%
Total % Variance explained	10%	10%	9%
Variance Partition Coefficient (ICC)	0.026	0.026	0.026

8. Discussion

8.1. The purpose of the project was to investigate using a value-add approach to better understand the factors that underlie the progress and performance of medical students. A data set including two cohorts of students starting medical school was available with a variety of potential predictors and outcome variables. For the purpose of the analysis the outcome variable with the best univariate relationship with the predictors was chosen but the relationships were disappointing being at best moderate in size. **The absence of an appropriate directly relevant outcome variable reflecting students' academic performance in medical school that differentiates between schools meant that the potentially strong relationship between prior attainment and university academic performance could not be modelled.** Table 8.1 lists alternate outcome variables that might be available for future research such as the forthcoming Medical Licensing Assessment (MLA)

8.2. The currently available outcome variable that was most suitable for a value-add analysis was the FPAS SJT. The SJT is designed to measure the professional attributes associated with being an F1 doctor such as Commitment to Professionalism, Coping with Pressure, Effective Communication, Patient Focus and Working Effectively as Part of a Team. These facets of performance are less related to cognitive ability than academic performance. Therefore, **the available predictors which reflect cognitive ability and academic attainment were not expected to explain a large amount of variance in SJT scores.**

8.3. The modelling showed that there was a statistically significant impact but relatively small amount of variance between medical schools on the SJT outcome measure. Intra class correlation had a value of 0.043 meaning that 4% of the variance in SJT scores was attributable to differences between medical schools. This is a relatively low level of attributed variation indicating there tends to be a similarity between medical school outcomes and this may make it harder to identify the moderator variables that might be related to this small difference.

8.4. The impact of prior attainment on the SJT scores, while not negligible was not as large as was hoped. The cognitive tests explained 6-7% of the total variance in SJT scores. One possible way to achieve a better overall goodness of fit would be to find a more cognitive-oriented medical school outcome variable to study. Such a measure would be likely better predicted by educational outcomes like A-levels at age 18 and the UKCAT tests. Although the prior attainment data are not strongly related to the SJT outcome variable at the individual level, it does account for around half of the apparent differences between medical schools in students SJT performance. That is half the school level differences in achievement on the SJT are accounted for by differences in the prior attainment levels of the students in the different medical schools. Of the remaining SJT variance very little (2%) is attributable to differences between medical schools. **This indicates that students' prior attainment is an important factor in explaining the observed raw differences between medical schools' average FY1 SJT scores.**

8.5. In relation to the equivalent findings from the UK school effectiveness literature, the overall fit of the "prior attainment only" model for medical schools is relatively poor indicated by the total variance explained (6-7%). In contrast, at the secondary school level the overall goodness of fit would typically be around 50% for GCSE and A-level outcomes (Thomas, 2001; Goldstein & Thomas, 1996; O'Donoghue et al., 1997). One explanation is that value-added methods may be more appropriate for statutory schooling than HEI sector, among other reasons due to fact that the variance in student outcomes and prior attainment is not so restricted by selection on entry. At the Higher Education level, very little work on value-added measures has been completed. One UK study reports relevant goodness of fit indicators for apparently similar analyses and it is encouraging that the findings at the lower end are fairly similar to the current study. Chapman (1996) examined the degree outcomes (proportion of first and upper second degrees) in eight subjects of study at the department level (Accountancy, Biology, Civil Engineering, French, History, Mathematics, Physics and Politics), over a 21-year period (1973-1993) in the UK. He found that the variation in the proportion of good degrees that was explained by entry qualifications was moderate or low, ranging from only 5% for Politics to 24% for Mathematics. Our findings are just below the lower end of this range.

8.6. Another possible explanation is the mismatch between the predictor variables which are cognitive in nature and the outcome variable which is a measure of contextual behaviour although it does have a cognitive element. Despite the low goodness of fit, the percentage of school level variance explained by prior attainment measures for the medical schools results (54-58%) are only slightly lower than the equivalent figures estimated for GCSE and A-level outcomes in secondary schools (70-82%) (Thomas, 2001; Goldstein & Thomas, 1996; O'Donoghue et al., 1997). This indicates there is explanatory power in the model at the level of the institution. This finding is also in line with McManus et al (2008) who examined the impact of entry qualifications to medical school on MRCP and Paces exams. They did not have access to individual level entry scores but only aggregate scores for institutions and found that prior attainment could account for around a half of the variance between schools. With regard to the intra-school correlation, this essentially represents the institutional effect (plus any other effects, either unmeasured or not controlled for in the analysis) and is interpreted as the percentage of total variance in student outcomes that can be attributable to differences between schools. Similar to the findings on the overall goodness of fit for the prior attainment only model, again this appears to be somewhat lower for medical schools (3%) than has previously been reported for secondary schools (8-10%) (Thomas, 2001; Goldstein & Thomas, 1996; O'Donoghue et al., 1997)..

8.7. In these preliminary analyses approximately one third of the 29 medical schools have value-added scores that are statistically significantly different from zero (i.e. not what would be expected given prior attainment). About half these are schools where the FY1 SJT performance is better than would be expected given the level of prior performance. It could be said that these schools are adding value by progressing students faster or further than would be expected. Other schools have a value-added score that is statistically significantly lower than would be expected given the students' scores on tests before entry. It should be remembered that these findings are tentative and need further exploration using a larger dataset and alternative medical school student outcome measures with a greater cognitive orientation.

8.8. The next stage of the analysis examined the impact of demographic and socioeconomic variables. These accounted for a further 3% of total variance in outcome scores although including them in the model resulted in a slight drop in explained school variance. The largest effects were seen for Gender and BME status which both accounted for between 20% and 25% of the explained variance in SJT scores. Women tended to perform better than men and White students had higher scores than those from BME backgrounds in the database.

8.9. The impact of socio-economic variables was much smaller. Only Free School meals and Income support reached statistical significance and the size of the effect was small <0.1 standard deviation. However there were additional significant findings when prior attainment was not taken into account. These included parents' occupation and education. Therefore the modelling shows that some group differences are confounded with and fully accounted for by differences in prior attainment for different groups. Importantly, for Free School Meals/Income Support this is not the case and there is an impact beyond that which is due to prior attainment levels.

8.10. The study has shown that, in principle, value-add methodology can be applied to medical school data. Despite limitations due to imperfect matching of the prior attainment variables and the outcome variables resulting in poor model fit, the analysis was able to account for half the between school variation and show that socio-economic and other WP variables can account for some of the variance in SJT outcome scores.

8.11. Gender, Age and BME status were all found to relate to FY1 SJT outcomes. Medical schools should therefore take particular care that their practices do not favour those from one or more of these groups, even if the effect is unintentional. Neither parent's education nor occupation showed a relationship with performance and neither did growing up in a low participation neighbourhood. **The only socio-economic indicator that showed a small relationship with the outcome variable was whether the individual had received free school meals or**

the family and received income support. These individuals tended to have marginally lower outcome scores on the SJT.

8.12. **We have also illustrated that a value-add methodology when applied in the medical context can potentially provide some useful results to understand the impact of prior attainment on progress and differential performance of medical schools. It can also provide a more sensitive analysis of the impact of demographic variables on performance.**

8.13. It is acknowledged that due to limitations of the data currently available, any substantive findings discussed here are tentative and need further confirmation using larger datasets and more student cohorts. Our primary intention in this study was to explore the value-add methodology as applied to medical schools. In future, it would also be preferable to build the data set based on outcome cohorts (all those who completed medical school in a particular year) rather than based on the starting year of candidates. Students completing in any given year across the UK will have started their studies across a range of three years as medical degrees vary in length from 4 to 6 years. The UKMED database has recently been extended from 2007 to 2014 and this will now allow the selection of a sample based on outcome year.

8.14. Employing a larger dataset would also be useful to explore the possibility of differential medical school effectiveness for different groups of students, even though the current findings suggest that this is unlikely to be found, given the small differences found overall for all students.

8.15. We have identified a number of issues in using HESA tariff data in this context which need to be resolved to ensure that results are robust. These include:

- Presence of low outliers in the distribution which do not represent levels of academic achievement consistent with capacity to undertake medical training. In this study these outliers are assumed to be incorrect and have been excluded.
- Presence of high outliers which represent values that are difficult to interpret
- Importance of using metrics which provide equivalent scores for qualifications from different sources. We saw differences between Scottish and other qualifications which were potential confounding factors with some approaches to quantifying secondary qualifications.

8.16. Further work could investigate potential process or moderator variables which might explain why some schools progress students better than others. Candidate variables would include issues related to the type of course (e.g. graduate entry, standard entry or medicine with a preliminary year Medical Schools Council (2016)), the structure of the degree, the way it is taught (e.g. problem-based-learning), structure of placements; issues relating to candidate selection (e.g., emphasis on academic versus non-academic selection criteria, ratio of public to state school educated, proportion of overseas educated students). Other potentially relevant moderator variables could be collected as part of the GMC's Medical Schools Annual Return (e.g. participation in test preparation activities) to help inform the design of future research studies.

8.17. For future use of the value-add approach it is critical to examine appropriate prior attainment and outcome measures that align more closely. To maximise the usefulness and interpretability of a value-add analysis Table 8.1 presents some potential alternative outcome variables that could be analysed.

Table 8.1 Alternate outcome variables

Variable	What it is	Why it might be useful	Potential Drawbacks	Availability
End of course academic outcomes	School assessments set on a common scale – but not based on ordinal measures such as deciles.	Provides a more academic criterion, but one which can vary across schools	Because schools approach assessment differently, difficult to compare scores across institutions. Schools may not be willing to release this data.	The UKCAT consortium has collected theory and skill scores for each student for each year (see McManus et al (2013). The UKCAT-12 study for a description) but only for UKCAT schools. The inclusion of these scores within UKMED is currently being negotiated.
Common content score from end of training exams	The schools have agreed to use some common elements in their assessments.	Scores based on these items would be equivalent	The common items represent a relatively short assessment that may not provide an accurate assessment.	These data are not held centrally by the Medical Schools Council
Prescribing Safety Assessment	New assessment that all students must pass to be deemed fit for clinical practice	Provides a common criterion across all schools	Not intended as an academic assessment. The purpose of the assessment is to ensure minimum competency and therefore it may not be designed to differentiate well among the higher performers.	Data from 2014 - 2016 have been loaded into UKMED.
Medical Licensing Exam	Where students are required to take a common exam in order to qualify for clinical practice.	Provides a common standard to evaluate learning during medical school.	The exam will be designed with practice safety in mind and may not produce the differentiation between students that is desirable – particularly at the top of the distribution.	MLA will be piloted with UK Graduates (Q2 2019 – Q2 2020)[5] and fully implemented in 2022.

[5]Council meeting, 2 June 2015 *Taking forward work on a UK licensing assessment* http://www.gmc-uk.org/10___Taking_forward_work_on_a_UK_licensing_assessment.pdf_61114454.pdf

Variable	What it is	Why it might be useful	Potential Drawbacks	Availability
Royal College Licensing Exams	Students take RC exams at the end of their speciality training.	The exams are the same for all trainees and so provide a common outcome standard to use as a criterion measure. The exams reflect successful qualification at speciality level which is a desirable outcome.	Substantial time lag between entering training and qualifying in specialism. Results will reflect practices that may no longer be current – e.g. medical school may have redesigned training since the cohort was trained.	All Royal College Exam data from 2014 onwards have been loaded into UKMED. Some exams such as the MRCP Part 1 are sat by foundation doctors (i.e. shortly after leaving medical school)
Current FY1 SJT variable, but use the UKCAT SJT on application as predictor	UKCAT now includes an SJT measure[6].	This is likely to be a better predictor of FY1 SJT scores as they are criterion matched.	As the UKCAT SJT is relatively new (introduced in 2014), practically it will be several years before those students will qualify from medical school and sit the FY1SJT	UKMED includes data on the UKCAT SJT scores for those applying for entry in 2013 onwards.

[6] http://www.ukcat.ac.uk/about-the-test/situational-judgement/

9. References

Blanden, J., & Gregg, P. (2004). Family Income and Educational Attainment: A Review of Approaches and Evidence for Britain.

Blanden, J., & Machin, S. (2004). Educational Inequality and the Expansion of UK Higher Education. *Scottish Journal of Political Economy*, *51*(2), 230–249. https://doi.org/10.1111/j.0036-9292.2004.00304.x

Bodger, O., Byrne, A., Evans, P. A., Rees, S., Jones, G., Cowell, C., … Williams, R. (2011). Graduate entry medicine: Selection criteria and student performance. *PLoS ONE*, *6*(11). https://doi.org/10.1371/journal.pone.0027161

Carneiro, P., & Heckman, J. J. (2002). The Evidence on Credit Constraints in Post-Secondary Schooling.

Chapman, K. (1996). Entry Qualifications, Degree Results and Value-added in UK Universities. *Oxford Review of Education*, *22*(3), 251–264. https://doi.org/10.1080/0305498960220301

Chowdry, H., Crawford, C., Dearden, L., Goodman, A., & Vignoles, A. (2013). Widening participation in higher education: Analysis using linked administrative data. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, *176*(2), 431–457. https://doi.org/10.1111/j.1467-985X.2012.01043.x

Crawford, C. (2014). Socio-economic differences in university outcomes in the UK: drop-out, degree completion and degree class, (W14/31).

Crawford, C. (2014). The link between secondary school characteristics and university participation and outcomes, (June).

Crawford, C., & Greaves, E. (2015). *Ethnic minorities substantially more likely to go to university than their White British peers*. Retrieved from http://www.ifs.org.uk/publications/8042

Do, P. C. T., Parry, J., Mathers, J., & Richardson, M. (2006). Monitoring the widening participation initiative for access to medical school: Are present measures sufficient? *Medical Education*, *40*(8), 750–758. https://doi.org/10.1111/j.1365-2929.2006.02535.x

Galindo-Rueda, F., Marcenaro-Gutierrez, O., & Vignoles, A. (2004). The Widening Socio-economic Gap in UK Higher Education.

Garlick, P., & Brown, G. (2008). Widening participation in medicine. *British Medical Journal*, *336*, 1111.

Glennerster, H., Hills, J., Kiernan, K., Le Grand, J., Power, A., & Propper, C. (2001). *Centre for Analysis of Social Exclusion*. Retrieved from http://sticerd.lse.ac.uk/Case

Goldstein, H., & Thomas, S. (1996). Using Examination Results as Indicators of School and College Performance. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *159*(1), 149–163.

Griffin, B., & Hu, W. (2015). The interaction of socio-economic status and gender in widening participation in medicine. *Medical Education*, *49*(1), 103–113. https://doi.org/10.1111/medu.12480

Haq, I., Higham, J., Morris, R., & Dacre, J. (2005). Effect of ethnicity and gender on performance in undergraduate medical examinations. *Medical Education*, *39*(11), 1126–1128. https://doi.org/10.1111/j.1365-2929.2005.02319.x

Health Informatics Centre (October 2015). *Safe Haven User Guide.* Available here: http://www.dundee.ac.uk/media/dundeewebsite/hic/documents/Safe%20haven%20User%20Guide.pdf. Accessed 4 September 2016

HEFCE. (2003). Schooling effects on higher education achievement, (July).

HEFCE. (2014). Differences in degree outcomes: Key findings, (March), 1–62. Available here:

http://www.hefce.ac.uk/media/hefce/content/pubs/2014/201403/HEFCE2014_03.pdf

http://www.isfp.org.uk/EPM/Pages/EPM.aspx Accessed 21st October 2016

https://newsite.hesa.ac.uk/support/documentation/tariffable. Accessed 17th October 2016.

http://www.theguardian.com/uk/2002/oct/01/highereducation.choosingadegree Accessed 21st October 2016.

http://www.ukcat.ac.uk/about-the-test/what-is-the-ukcat/ Accessed 17th October 2016.

James, D., Yates, J., & Nicholson, S. (2010). Comparison of A-level and UKCAT performance in students applying to UK medical and dental schools in 2006: cohort study. *BMJ (Clinical Research Ed.)*, *340*.

Lievens, F., Patterson, F., Corstjens, J., Martin, S., & Nicholson, S. (2016). Widening access in selection using situational judgement tests: Evidence from UKCAT. *Medical Education*, *50*(6), 624–636.

Machin, S., & Vignoles, A. (2004). Educational Inequality: The Widening Socio-Economic Gap. *Fiscal Studies*, *25*(2), 107–128.

McManus, I. C., Elder, A. T., De Champlain, A., Dacre, J. E., Mollon, J., & Chis, L. (2008). Graduates of different UK medical schools show substantial differences in performance on MRCP(UK) Part 1, Part 2 and PACES examinations. BMC Medicine, 6:5.

McManus, I. C., Dewberry, C., Nicholson, S., & Dowell, J. S. (2013). The UKCAT-12 study: educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a cross-sectional collaborative study of 12 UK medical schools. *BMC Medicine*, *11*, 1–25. https://doi.org/10.1186/1741-7015-11-244

Medical Schools Council. (2011). Educational Performance Measure (EPM) Framework, 1–3.

Medical Schools Council. (2016). *Implementing Selecting for Excellence: a Progress Update*. Retrieved from http://www.medschools.ac.uk/SiteCollectionDocuments/Selecting-for-Excellence-2016-update-MSC.pdf

Medical Schools Council (2016). *Entry requirements for UK medical schools 2017 entry*. Retrieved from http://www.medschools.ac.uk/SiteCollectionDocuments/MSC-Entry-requirements-for-UK-medical-schools.pdf

Munoz-Chereau, B. & Thomas, S. M. (2016) Educational effectiveness in Chilean secondary education: Comparing different 'value added' approaches to evaluate schools. Assessment in Education. 23 (1) pp 26-52 http://dx.doi.org/10.1080/0969594X.2015.1066307

National Statistics Socio-economic Classification: http://www.ons.gov.uk/ons/guide-method/classifications/archived-standard-classifications/soc-and-sec-archive/the-national-statistics-socio-economic-classification--user-manual.pdf. Accessed 21st October 2016.

O'Donoghue, C., Thomas, S., Goldstein, H., & Knight, T. (1997). *1996 DfEE study of Value Added for 16-18 year olds in England*. Retrieved from http://www.bristol.ac.uk/cmm/team/hg/alevdfee.pdf

Ogg, T., Zimdars, A., & Heath, A. (2009). Schooling effects on degree performance: a comparison of the predictive validity of aptitude testing and secondary school grades at Oxford University. *British Educational Research Journal*, *35*(5), 781–807. https://doi.org/10.1080/01411920903165611

Raffe, D. (2007). Making Haste Slowly: the evolution of a unified qualifications framework in Scotland. *European Journal of Education*, *42*(4), 485–502. https://doi.org/10.1111/j.1465-3435.2007.00322.x

Searcy, C. A., Dowd, K. W., Hughes, M. G., Baldwin, S., & Pigg, T. (2015). Association of MCAT scores obtained with standard vs extra administration time with medical school admission, medical student performance, and time to graduation. *Jama*, *313*(22), 2253–2262. https://doi.org/10.1001/jama.2015.5511

Secretary of State for Education and Skills. (2004). *Department for Education and Skills: Five Year Strategy for Children and Learners*.

Seyan, K., Greenhalgh, T., & Dorling, D. (2004). The standardised admission ratio for measuring widening participation in medical schools: analysis of UK medical school admissions by ethnicity, socioeconomic status, and sex. *BMJ (Clinical Research Ed.)*, *328*(February 2009), 1545–1546. https://doi.org/10.1136/bmj.328.7455.1545

Simmenroth-Nayda, A., & Görlich, Y. (2015). Medical school admission test: advantages for students whose parents are medical doctors? *BMC Medical Education*, *15*(1), 81. https://doi.org/10.1186/s12909-015-0354-x

Smith, J., & Naylor, R. (2002). *Schooling effects on subsequent university performance: Evidence for the UK university population.Warwick Economic Research Papers No.657*.

Steven, K., Dowell, J., Jackson, C., & Guthrie, B. (2016). Fair access to medicine? Retrospective analysis of UK medical schools application data 2009-2012 using three measures of socioeconomic status. *BMC Medical Education*, *16*(1), 11. https://doi.org/10.1186/s12909-016-0536-1

Taylor, C. a, Green, K. E., & Spruce, A. (2015). Evaluation of the effect of socio-economic status on performance in a Multiple Mini Interview for admission to medical school. *Medical Teacher*, *37*(1), 59–63. https://doi.org/10.3109/0142159X.2014.923562

Thomas, S. (2001): Dimensions of Secondary School Effectiveness: Comparative Analyses Across Regions, School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 12:3, 285-322

Thomas, SM 2010, 'Assessment and the evaluation of institutional effectiveness'. in B McGaw, P Peterson & E Baker (eds), International Encyclopedia of Education (3rd edition). Elsevier.

Thomas, S., & Mortimore, P. (1996). Comparison of value-added models for secondary-school effectiveness. *Research Papers in Education*, *11*(1), 5–33. https://doi.org/10.1080/0267152960110103

Thomas et al. (2013). Evidence of CVA School Effects in China and UK. Symposium Using Contextualised Value Added to examine school effectiveness as a tool for accountability. In *ICSEI Annual Conference*. Santiago.

Tiffin, P. A., McLachlan, J. C., Webster, L., & Nicholson, S. (2014). Comparison of the sensitivity of the UKCAT and A-levels to sociodemographic characteristics: a national study. *BMC Medical Education*, *14*, 7. https://doi.org/10.1186/1472-6920-14-7

UCAS Analysis and Research. (2016a). *End of Cycle 2015 Data Resources: DR3_015_03.*

UCAS Analysis and Research. (2016b). *End of Cycle 2015 Resources: DR3_015_01.* Retrieved from End of Cycle 2015 Data Resources

10. Appendix A: Search Strategy for Literature Review

10.1. **Search terms**: widening participation medici*; widening participation healthcare; widening participation selection; widening access medici*; widening access healthcare; widening access selection; socio-economic medici*; school type medici*; parental education medici*; disability medici*; income support medici*; free school meals medici*; postcode medici*; POLAR; ethnicity medici*

10.2. **Databases searched:** PubMed, Google Scholar, Web of Science, Advances in Health Sciences Education

10.3. **Inclusion criteria:** published studies, research reports and chapters which analyse original or secondary data; qualitative, quantitative and mixed-methods designs; studies or chapters set in the context of medical or other healthcare professions; studies focussing on selection and in-training performance; research articles and reports examining prior attainment in context of demographic variables; additional studies shared by subject matter experts in Work Psychology Group's project team; research into how demographic variables impact on entry to, and performance during higher education degrees

10.4. **Exclusion criteria:** studies published before the year 2000; opinion or commentary pieces that do not report data, unpublished articles and chapters; studies which do not include analysis of scores at selection or during training

11. Appendix B: Sensitivity Analysis: Comparison of Methods for Scoring Educational Qualifications

11.1. This section reports the impact of using different methods for calculating a prior attainment score based on educational qualifications.

11.2. HESA tariff scores ranged from 0 to over 1000. A score of 300 can be attained by achieving a B grade in three A-levels. It is unlikely that someone with this level of qualification would be offered a university place to study medicine. Therefore scores of 300 and below were excluded. This variable is labelled z_HESA trunc in table 11.1 below.

11.3. Very high HESA scores may also be anomalous since most students do not have the opportunity to sit sufficient examinations or subjects to achieve scores as high as 900 or 1000. Therefore we considered capping the HESA score at 800 with all scores above this value set to 800. The result is shown in the second section of table 11.1 below. There is no improvement in prediction through capping high scores in this manner and so this approach was not used.

11.4. The data set included both scores for individual examinations (e.g. A-level subject results) and the total score when the tariff system is applied. Differences in total tariff scores will reflect candidate ability but are also impacted by the opportunities a candidate had. For example candidates who attended a school which supported them in preparing for, and allow them to sit more A-levels will be likely to have higher tariff scores than those who attended schools with a more restrictive A-level programme. McManus et al (2013) proposed using a score based on the three best A-level scores for candidates (excluding General Studies). In their study of the relative validity of the UKCAT and educational achievement in 12 schools for predicting first year results they applied this approach to both A-levels and Scottish Highers. The full methodology is described in the technical report of the project[7]. We refer to this approach as the UKCAT-12 approach.

11.5. The UKCAT-12 approach would be preferred if it resulted in better prediction of outcome results as McManus et al found. The third part of table 11.1 shows this approach to lead to a stronger relationship with SJT scores. However there are drawbacks for the current study as it relies on standard educational qualifications while those in a widening access group may well have qualified through an alternative route. Given the focus on widening access it is undesirable to drop all candidates without standard school leaver qualifications from the sample.

11.6. Another difficulty in creating a score representing educational qualifications is to find an approach which provides equivalent scores from different qualification regimes. While the majority of the sample had A-Level qualifications, a substantial minority had Scottish qualifications with scores on Highers. The UKCAT12 study deals with this difficulty by standardising within group, however this has the potential to remove variance of interest for this study. We are interested in comparing medical schools, but Scottish schools will predominantly have scores derived from Highers and other schools scores from A-levels. Combining within group z scores makes the assumption that the two groups have the same distribution of ability. This may not be the case, for example, there may be differential self-selection factors in applicants to medical schools in Scotland and elsewhere.

[7] The UKCAT-12 Study, Technical Report: Educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in twelve UK medical schools (2012) IC McManus, C Dewberry, S Nicholson, J Dowell

Table 11.1 Comparing the prediction of the Z - SJT score from the available measures of educational attainment.

N = 6,978	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	0.176	0.010		18.495	.000
z_HESA trunc > 300	0.079	0.010	0.099	8.324	.000
R^2 (adj) = 0.010					

N = 6,978	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	0.177	0.010		18.510	.000
z_HESA_Trunc > 300 and < 800	0.080	0.010	0.099	8.337	.000
R^2 (adj) = 0.010					

N = 6,879	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	0.132	0.011		12.184	.000
Z Score A-level or SQA	0.227	0.025	0.110	9.176	.000
R^2 (adj) = 0.012					

11.7. One sign that this may be the case was seen in the pattern of correlations for the UKCAT12 score. For example, Table 11.2 shows that the pattern of correlations is a little different for the A-level and SQA qualifications. A regression analysis found an adjusted R^2 of 0.012 for the UKCAT12 methodology scores as predictors of the SJT. When calculated separately for A-level and SQA qualifications the values were 0.009 and 0.059. These are differences of an order of magnitude. No similar impact is seen for the HESA tariff scores where the equivalent values are 0.091 and 0.125. The latter pattern suggests a more consistent combination of scores.

11.8. Because of the unexplained differences in results for A-level and SQA based qualifications, and because using the Tariff approach allowed the inclusion of more individuals, the Tariff approach was chosen for this study.

Table 11.2 Comparing the prediction of the Z - SJT score from educational attainment in A-level and SQA samples.

N = 6,205 (A-level only)	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	0.171	0.010		16.816	.000
z_HESA trunc > 300	0.075	0.010	0.091	7.225	.000
R^2 (adj) = 0.008					

N=674 (SQA only)	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	0.205	0.031		11.571	.000
z_HESA trunc > 300	0.102	0.031	0.125	7.455	.000
R^2 (adj) = 0.014					

N = 6,205 (A-level only)	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	0.132	0.011		11.571	.000
Z Score A-level	0.197	0.026	0.094	7.455	.000
R^2 (adj) = 0.009					

N=674 (SQA only)	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	0.126	0.033		3.832	.000
Z Score SQA	0.454	0.069	0.247	6.597	.000
R^2 (adj) = 0.059					

12. Appendix C: Sensitivity Analysis: Comparison of Separate versus Combined UKCAT Scores

12.1. Prediction of the SJT score was improved by using the UKCAT subscales instead of the total score. As the tolerance suggested that there was no problem with multi-collinearity with the subscale score these were used in the model. Table 12.1 compares the two regression models. Table 12.2 shows the correlations between SJT, EPM and the different methods for calculating educational attainment.

Table 12.1 Comparing the prediction of SJT score from UKCAT subscale scores or the total score. N = **6,978**

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
(Constant)	0.166	0.009		17.930	.000	.148	.185		
Zscore: Abstract Reasoning	0.048	0.010	.059	4.774	.000	.028	.068	0.877	1.140
Zscore: Decision Analysis	0.055	0.010	.068	5.417	.000	.035	.075	0.852	1.174
Zscore: Quantitative Reasoning	0.030	0.010	.036	2.887	.004	.010	.050	0.854	1.171
Zscore: Verbal Reasoning	0.163	0.010	.197	15.901	.000	.143	.183	0.866	1.154
R^2 (adj) = 0.068									

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	0.165	.009		17.673	.000	0.147	0.183
z_ukcat_total	0.199	.010	0.239	20.577	.000	0.180	0.218
R^2 (adj) = 0.057							

Table 12.2 Correlations between methods of calculating educational attainment, EPM and SJT scores. N = **6,978**

	Z_SJT	FPAS_EPM_Final	z_ukcat_total	z_HESA trunc > 300	z_HESA_Trunc > 300 and < 800	z_Alevel		z_
FPAS_EPM_Final N =6,978	0.288**							
z_ukcat_total N =6,978	0.239**	0.139**						
z_HESA trunc > 300 N =6,978	0.099**	0.130**	0.271**					
z_HESA_Trunc > 300 and < 800 N =6,978	0.099**	0.132**	0.273**	0.998**				
Z Score A-level N =6,237	0.051**	0.127**	0.157**	0.257**	0.258**			
Z SQA N = 680	0.211**	0.197**	0.241**	0.336**	0.346**	-0.568**		
Z Score A-level or SQA N =6,879	0.110**	0.171**	0.230**	0.380**	0.382**	0.742**		0.

13. Appendix D: Sensitivity Analysis: Imputing Missing Values

13.1. Imputing missing values using the SPSS procedure and predicting the SJT score from the UKCAT sub scales and the HESA Tariff using all cases with an SJT score gave a very slight improvement from 0.072 to 0.075, Table 13.1 gives the overall model and show the change in R^2 from the original data to the model using imputed missing values. Table 13.2 gives the regressions with the original data and the imputed versions. As the improvement in R^2 was small the original data were used.

Table 13.1 Predicting Z –SJT using imputed data compared to the original data – model summary. N= 9,511 for imputed and N = 7,620 for original data

Imputation Number	R	R^2 Square	Adjusted R^2	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
Original data	0.269	0.073	0.072	0.78173	0.073	119.122	5	7614	0
1	0.274	0.075	0.075	0.80079	0.075	154.25	5	9,505	0
2	0.276	0.076	0.075	0.8004	0.076	156.255	5	9,505	0
3	0.275	0.076	0.075	0.80055	0.076	155.49	5	9,505	0
4	0.275	0.076	0.075	0.80054	0.076	155.522	5	9,505	0
5	0.271	0.074	0.073	0.80142	0.074	151.026	5	9,505	0

Table 13.2 Predicting Z –SJT using imputed data compared to the original data

Imputation Number		Unstandardized Coefficients		Standard-ized Coeffic-ients	t	Sig.	Fraction Missing Info.	Relative Increase Variance	Relative Efficiency
		B	Std. Error	Beta					
Original data	(Constant)	.147	.009		16.314	0.000			
	z_HESA	0.025	0.009	0.031	2.737	0.006			
	Z - Abstract Reasoning	0.051	0.01	0.061	5.200	0.000			
	Z - Decision Analysis	0.053	0.01	0.064	5.344	0.000			
	Z – Quantitative Reasoning	0.029	0.01	0.034	2.834	0.005			
	Z – Verbal Reasoning	0.163	0.01	0.196	16.460	0.000			
1	(Constant)	0.13	0.008		15.746	0.000			
	z_HESA	0.044	0.008	0.058	5.623	0.000			
	Z - Abstract Reasoning	0.052	0.008	0.066	6.442	0.000			
	Z - Decision Analysis	0.052	0.009	0.062	5.884	0.000			
	Z – Quantitative Reasoning	0.032	0.009	0.038	3.513	0.000			
	Z – Verbal Reasoning	0.161	0.009	0.193	18.158	0.000			
2	(Constant)	0.13	0.008		15.808	0.000			
	z_HESA	0.051	0.008	0.065	6.372	0.000			
	Z - Abstract Reasoning	0.063	0.008	0.081	7.993	0.000			
	Z - Decision Analysis	0.046	0.009	0.055	5.175	0.000			
	Z – Quantitative Reasoning	0.024	0.009	0.029	2.679	0.007			
	Z – Verbal Reasoning	0.16	0.009	0.192	18.103	0.000			
3	(Constant)	0.13	0.008		15.768	0.000			
	z_HESA	0.05	0.008	0.065	6.310	0.000			
	Z - Abstract Reasoning	0.054	0.008	0.069	6.778	0.000			
	Z - Decision Analysis	0.052	0.009	0.062	5.894	0.000			
	Z – Quantitative Reasoning	0.026	0.009	0.031	2.926	0.003			
	Z – Verbal Reasoning	0.16	0.009	0.191	17.964	0.000			

Imputation Number		Unstandardized Coefficients		Standard-ized Coeffic-ients	t	Sig.	Fraction Missing Info.	Relative Increase Variance	Relative Efficiency
4	(Constant)	0.129	0.008		15.655	0.000			
	z_HESA	0.047	0.008	0.06	5.812	0.000			
	Z - Abstract Reasoning	0.054	0.008	0.07	6.945	0.000			
	Z - Decision Analysis	0.056	0.009	0.066	6.297	0.000			
	Z – Quantitative Reasoning	0.025	0.009	0.03	2.768	0.006			
	Z – Verbal Reasoning	0.161	0.009	0.194	18.223	0.000			
5	(Constant)	0.128	0.008		15.585	0.000			
	z_HESA	0.04	0.008	0.051	4.960	0.000			
	Z - Abstract Reasoning	0.057	0.008	0.073	7.194	0.000			
	Z - Decision Analysis	0.053	0.009	0.063	5.978	0.000			
	Z – Quantitative Reasoning	0.03	0.009	0.035	3.308	0.001			
	Z – Verbal Reasoning	0.158	0.009	0.19	17.828	0.000			
Pooled	(Constant)	0.129	0.008		15.639	0.000	.009	.009	.998
	z_HESA	0.046	0.009		4.964	0.000	.296	.372	.944
	Z - Abstract Reasoning	0.056	0.009		5.993	0.000	.308	.391	.942
	Z - Decision Analysis	0.052	0.01		5.340	0.000	.176	.198	.966
	Z – Quantitative Reasoning	0.027	0.01		2.828	0.005	.141	.154	.972
	Z – Verbal Reasoning	0.16	0.009		17.853	0.000	.022	.023	.996

14. Appendix E: Sensitivity Analysis: Curvilinear Models

14.1 The relationship between the predictors and outcome variables may be non-linear rather than linear as the models tested assume. One method of identifying curvilinear relationships is to include squared and cubed terms for the predictor variables in the equation. This approach allows the modelling of the most likely curvilinear models. If the model fit improves with the inclusion of these terms it is likely that the underlying relationship is not linear.

14.2 Table 14.1 shows that there is minimal improvement in the model with the addition of the additional squared and cubed variables. Thus no curvilinear effects have been identified.

Table 14.1 Prior attainment model for 2 samples with squared and cubed terms

Model	n=4691	n=6978
-2*log likelihood	10883.7.8	16073.0
Reduction in -2*log likelihood	21.1	28.1
Intercept (se)	0.214 (0.029)	0.196 (0.026)
Parameter HESA Tariff	0.004 (0.014)	0.011 (0.012)
Square HESA Tariff	-0.014 (0.011)	-.009 (0.009)
Cubed HESA Tariff	0.004 (0.003)	0.003(0.002)
UKCAT Abstract Reasoning	0.080 (0.018)	0.083 (0.015)
Square UKCAT AR	-0.000 (0.007)	-0.001 (0.006)
Cubed UKCAT AR	**-0.010 (0.003)**	**-0.010 (0.003)**
UKCAT Decision Analysis	0.008 (0.019)	0.030 (0.015)
Square UKCAT DA	0.006 (0.007)	-0.003 (0.006)
Cubed UKCAT DA	0.008 (0.004)	0.005 (0.003)
UKCAT Quantitative Reasoning	0.026 (0.019)	0.030 (0.015)
Square UKCAT QR	-0.008 (0.007)	-0.004 (0.006)
Cubed UKCAT QR	-0.000 (0.003)	-0.003 (0.003)
UKCAT Verbal Reasoning	0.153 (0.019)	0.152 (0.015)
Square UKCAT VR	-**0.018 (0.008)**	**-0.019 (0.006)**
Cubed UKCAT VR	-0.003 (0.004)	-0.001 (0.003
Variance (SE) Med School	0.013 (0.005)	0.012 (0.004)
Residual Variance (SE) student	0.591 (0.012)	0.582 (0.010)
Total Variance	0.604	0.594
Med School % Variance explained	54%	54%
Total % Variance explained	7%	7%
ICC	0.022	0.020

15. Appendix F: FY1 SJT - Overview

Context:

15.1. Every year, nearly 8,000 final year medical students apply for junior doctor posts in the two-year Foundation Programme, which is a requirement for all medical graduates wishing to work as doctors in the UK. Competition into the programme is intensifying due to the expansion of UK medical schools and the ever-increasing number of international applications. The Department of Health recommended that an SJT was implemented to assess professional attributes, judgement and employability for a Foundation Programme post and to replace the open-ended competency based application questions previously used.

Test format and content:

15.2. The FY1 SJT presents applicants with scenarios they are likely to encounter as a Foundation Year 1 (F1) doctor and asks how they would react in these situations. Their responses are scored against a pre-determined key defined by subject-matter experts. The SJT ensures candidates selected have the aptitude and values required of a successful doctor. In order to understand and define the attributes required to be successful in the role, a job analysis of the F1 doctor role was undertaken. A person specification was developed based on this analysis. Each year, educational supervisors, clinical supervisors and other Foundation Programme experts contribute to the development of new test questions based on the person specification. These questions are reviewed, including input from Foundation doctors to ensure that the hypothetical situations are realistic and appropriate. From the job analysis, five target domains were identified to be assessed in the SJT and each item is designed to measure one of these. The five target domains are outlined in Table 15.1 along with examples of the kinds of scenarios which might be included under each.

Table 15.1. FY1 SJT target domains and example SJT Scenarios

Domains and Example SJT Scenarios	
Commitment to professionalism	*E.g. Issues of confidentiality such as hearing a colleague talking about a patient outside of work*
Coping with pressure	*E.g. Dealing with confrontation such as an angry relative*
Effective communication	*E.g. Gathering information and communicating intentions to other colleagues*
Patient focus	*E.g. Taking into account a patient's views/concerns*
Working effectively as part of a team	*E.g. Recognising and valuing the skills and knowledge of colleagues, when faced with a disagreement about a patient's care*

Outcomes and evaluation:

15.3. The SJT was piloted as part of the FPAS process in 2012 and has been in operational use since 2013, with results informing the allocation of medical school graduates to foundation programme places alongside a measure of educational attainment. Each year the SJT is subject to full psychometric analysis and results consistently show that the SJT is a reliable, valid and appropriate method for foundation selection. Applicant reactions to the test have been positive, with the majority of students indicating that the content of the test seems fair, relevant to the Foundation Programme, and appropriate for their level. A predictive validity study undertaken in 2015, compared performance on the test with performance on the foundation programme finding that the SJT demonstrated good predictive validity of supervisor-rated performance and incidence of remedial action[8].

Coaching and SJTs

15.4. SJTs can be designed to be less susceptible to coaching by tailoring their content and the response formats used and instructions given. The study by Lievens et al[9] of an SJT for Belgian medical school admissions is welcome as a propensity scoring analysis resolves problems associated with self-selection in coaching studies. Here, results showed coaching effects in the form of a 0.5 standard deviation improvement; however, the SJT design specification differed significantly from others. The FY1 SJT uses a more complex response format that employs two types of response ('Rank all five possible responses in order' and 'Choose the three best responses from a list of eight') that are significantly more cognitively loaded than those used in the response format in the Belgian test ('Pick the best response from a choice of four'). Lievens et al suggested that using complex, cognitively loaded formats is likely to make SJTs less prone to coaching effects

15.5. SJTs of more heterogeneous content (i.e. capturing a variety of domains) are less susceptible to coaching as they increase 'the range and specificity of strategies that must be learned and memorised by trainees'. The Belgian SJT targeted two domains (building relationships and communicating information), using one response format, across 30 items. By contrast, the FY1 SJT targets five domains, using two response formats, across 70 items. Other steps to reduce coaching effects include building complexity into scenarios so that candidates must engage with the question rather than using a simple response strategy, and maintaining a large, continually renewed item bank. Research shows that SJTs are less susceptible to coaching effects when they are constructed from experts' judgements or empirical keying rather than rules. Adopting cognitively oriented response instruction (i.e. 'What should you do?') rather than a behaviourally oriented format (i.e. 'What would you do?') makes SJTs less susceptible to self-deception and impression management[10] [11]and therefore to coaching.

15.6. Research is still required to examine whether SJT coaching produces genuine or artificial effects. Formal education and training in important domains (such as communication) could and, indeed, should be beneficial to the development of learners in the longer term. By contrast, coaching is usually short-lived and geared towards test-taking strategies. External providers (usually commercial) provide tips that are specifically intended to help a candidate 'pass' the SJT, rather than to facilitate detailed understanding about what constitutes effective behaviour in job-relevant situations. However, research is required to explore these differential effects in SJTs (skills development versus test-taking strategies) and to examine whether coaching is actually linked to training and job outcomes. SJTs measure understanding of effective behaviour in a given situation, in relation to non-academic attributes, such as empathy, integrity and teamwork, depending on the test specification. SJTs do not measure personality traits per se; they measure implicit trait policies (ITPs) and general experience (and, depending on level, specific job knowledge).[12] ITPs are beliefs about the costs and benefits of expressing certain traits, such as knowing that being agreeable is likely to be better in many situations. Higher-order interactions between individual differences in cognition, intellect, personality and affect, and links to training outcomes and job performance, represent an exciting area for SJT research.

---

[8] Cousans, F., Patterson, F., Edwards, H., Walker, K., McLachlan, J., & Good, D. (2017). Evaluating the complementary roles of an SJT academic assessment for entry into clinical practice. *Advances in Health Science Education*. DOI 10.1007/s10459-017-9755-4

[9] Lievens F, Buyse T, Sackett PR, Connelly BS. The effects of coaching on situational judgement tests in high-stakes selection. International *Journal of Selection and Assessment* 2012;20 (3):272–82.

[10] McDaniel M, Hartman N, Whetzel D, Grubb W. Situational judgement tests, response instructions, and validity: a meta-analysis. *Personnel Psychology* 2007;60 (1):63–91.

[11] Nguyen NT, Biderman MD, McDaniel M. Effects of response instructions on faking a situational judgement test. *International Journal of Selection and Assessment* 2005;13 (4):250–60.

[12] Motowidlo SJ, Beier ME. Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a situational judgement test. *Journal of Applied Psychology* 2010;95 (2):321–33.