Leading education and social research
Institute of Education
University of London

# Does an aptitude test affect socioeconomic and gender gaps in attendance at an elite university?

**Jake Anders**

## Department of Quantitative Social Science

Working Paper No. 14-07
April 2014

## Disclaimer



Department of Quantitative Social Science, Institute of Education, University of London
20 Bedford Way, London WC1H 0AL, UK

# Does an aptitude test affect socioeconomic and gender gaps in attendance at an elite university?


## Jake Anders[1]



### Abstract

The increasing use of aptitude tests as part of the admissions processes at elite English universities potentially has significant implications for fair access to these institutions. I attempt to isolate the impact of the introduction of one such test on the proportion of successful applicants by school type (as a proxy for socioeconomic status) and by gender using a difference in differences approach and administrative data from the University of Oxford. The introduction of the test coincided with the implementation of a guideline number of interviews per available place, significantly reducing the proportion of applicants offered an interview (by 14 percentage points) and, hence, increasing the proportion of interviewees offered places (by 3.6 percentage points). By gender, I find some evidence that these changes may be having differing effects at different stages of the admissions process, but not on each group's overall chances of securing an offer. I do not find any evidence that the policy has negative side effects on the chances of applicants from less advantaged socioeconomic backgrounds at any stage of the process.



**JEL classification:** I23, I 24

**Keywords:** Higher Education, Aptitude Test, Gender, Socioeconomic Gradient, Difference in Differences.



[1] Department of Quantitative Social Science, Institute of Education, University of London (jake@jakeanders.co.uk)



Acknowledgements

Thanks to my supervisors Lorraine Dearden and John Micklewright for their patient support. Thanks also to John Jerrim for helpful feedback and Claire Crawford for providing analysis to validate the school-area IDACI measure. Many thanks to Ian Crawford and Margaret Stevens of the Department of Economics, University of Oxford for clarifications and advice. Likewise to Elizabeth Frazer and members of the Admissions Office at New College, Oxford for their assistance with this project. I am very grateful to the University of Oxford Admissions Office for permission to use the data. Nevertheless, these individuals and organisations are not responsible for any errors made in the author's use of the data. Any views expressed are the author's own and do not necessarily reflect those of New College, Oxford or the University of Oxford. I gratefully acknowledge my studentship funded by the IoE as part of the ESRC NCRM ADMIN node.

# 1 Introduction

The increasing use of aptitude tests as part of the admissions processes at elite universities in Britain potentially has significant implications for fair access to these institutions. While the intention is to improve the efficiency of the process, making it easier to select individuals with a better 'aptitude' [1] for their university course, is this efficiency gain traded off against other aims of the admissions process? In particular, previous research suggests there are reasons to think aptitude testing may have side effects on the proportion of applicants from different socioeconomic backgrounds (Rothstein, 2002) and different genders (Tannenbaum, 2012) who get a place.

To explain this concern, let us take the example of fair access by socioeconomic status. There are at least two potential reasons that the introduction of an aptitude test could result in a smaller intake of those from lower socioeconomic backgrounds. First, the outcomes of the test could reflect skills acquired in previous education, hence skewing the distribution of those offered a place towards those who received certain kinds of schooling, or reflect training to the test, both of which might be of concern (Stringer, 2008). Alternatively, it could reflect genuine differences in aptitude for the university's degree programme across the socioeconomic spectrum. However, there are also reasons to see the possibility of the opposite effect as a result of the introduction of an aptitude test, with more offers of places made to those from less advantaged backgrounds. If more weight is given to aptitude test results over and above school examination results then this could help overcome bias in those indicators caused by schooling rather than underlying ability. This paper aims to identify which, if either, of these effects seems to dominate and hence understand the wider impact of using aptitude tests as a selection tool.

In 2007-2008, the University of Oxford, an elite British university, introduced an aptitude test as part of the admissions process for Economics-related subjects. I use administrative data from the University's admissions system, covering all undergraduate applications, to estimate the differential impact of the introduction of this test on applicants by their socioeconomic backgrounds and their gender. I employ a difference in differences framework: this attempts to control for any general trends using those seen in subjects where the aptitude test was not

[1]'Aptitude' is taken broadly as a measure of potential attainment, as against prior attainment such as measured by A Levels or GCSEs, or innate ability.

introduced, hence isolating the impact due to the policy change.

The paper proceeds as follows. In Section 2, I survey the literature on access to elite universities, identify important details about the use of aptitude tests in university admissions, and lay out the research questions for this paper. I then detail the admissions process at the University in Section 3 and describe the data used in this work in Section 4. Section 5 describes the changes in admissions during this period and identifies the particular features of the change in policy. It then lays out the empirical strategy for identifying the changes in outcomes that seem to be associated with its introduction and presents simple estimates of impact. I extend this using regression analysis, describing my models in Section 6 and presenting the results in Section 7. I consider an alternative way of looking at the results in Section 8 and conduct various robustness checks in Section 9, before concluding in Section 10.

# 2 Previous research and research questions

Why take an interest in the admissions processes of elite universities, and the introduction of an aptitude test in particular? I consider these questions in turn.

Given the higher wage premiums graduates from elite universities seem to command (Chevalier and Conlon, 2003), fair access to these institutions is important to future equality of opportunity. Furthermore, we cannot necessarily rely on insights about fair access to all universities to understand inequalities at elite universities; Pallais argues that "it is entirely plausible that barriers to enrollment at the most selective institutions are somewhat different than at the margin of enrollment" (Pallais and Turner, 2008, p.132) and as such the correct policy response may well be different.

The current UK government's belief is that "progress over the last few years in securing fair access to the most selective universities has been inadequate, and that much more determined action now needs to be taken" (Willetts, 2011). Previous research from both the UK and the US has highlighted concern about the equality of opportunity in access to elite Higher Education institutions. In Anders (2012a) I showed that, among young English people who do attend university, those from the bottom income quintile group are almost 20 percentage points less likely to attend a Russell Group institution (a group of elite UK universities) than those from the top income quintile group. Similarly, analysis by Boliver (2013) highlighted that Russell Group applicants from state schools are less likely to receive offers of admission from Russell Group universities in comparison with their equivalently qualified peers from private schools. Such concerns also exist in the US: "Less than 11 percent of first–year students matriculating at 20 highly selective institutions were from the bottom income quartile of the income distribution" (Pallais and Turner, 2006, p.357).

Specifically regarding the University of Oxford, Bhattacharya et al. (2012) use administrative data from one undergraduate programme to estimate the expected performance of the marginal admitted candidate by sex and school type, arguing that in an academically fair process this threshold for admission would be equal between such groups. However, they estimate that the expected performance of the marginal candidate from an independent school is approximately 0.3 standard deviations higher than their state school counterpart. Similarly, the expected performance of the marginally admitted male candidate is about 0.6 standard deviations higher than their female counterpart.

Aptitude testing has become a much more important issue in recent years. As more students have begun to reach the upper bound of performance in A Levels (examinations taken by most English students aiming for entry to Higher Education, usually at age 18) it has become harder for universities to differentiate between potential students at the top end of the ability distribution[2]. This has led to an increasing use of aptitude tests among elite institutions, including the BioMedical Aptitude Test and United Kingdom Clinical Aptitude Test for admission to medical courses at many universities; the Physics Aptitude Test, at the University of Oxford; and, the focus of this paper, the Thinking Skills Assessment at the University of Oxford, the University of Cambridge and University College London (Admissions Testing Service, 2013b);. However, an important question is whether this response is a sensible course of action, especially in the light of the inequalities discussed above.

If we see aptitude as a measure of potential ability in a given field, then aptitude tests should be effective at predicting the performance of candidates once they reach university and should do so without being biased by candidates' other characteristics. Unfortunately, McDonald et al. (2001b) find little evidence that the Scholastic Aptitude Test (SAT) predicts attainment once at college in the US any better than high school record alone. These findings were replicated in a pilot study in Britain (McDonald et al., 2001a). A more recent Department of Business, Inno-

[2]This analysis covers the period before the introduction of the new A* grade for A-Levels, which has ameliorated this problem to some extent.

vation and Skills (BIS) report comes to similar conclusions, arguing that the SAT does not provide significantly more information on applicants' likely performance at undergraduate level, relative to a baseline of GCSE (English school examinations taken at the end of compulsory education) attainment scores (Kirkup et al., 2010, p.20).

On the question of bias in aptitude test scores, the fact that "low-income students not only are less likely to take college placement tests but also tend to have lower scores on these exams" (Pallais and Turner, 2008, p.135) suggests, on the face of it, that aptitude testing could cause more harm than good. In addition, Pallais and Turner (2008) note that the "gap [in aptitude tests between low and high income students] is particularly marked at the top of the distribution from which elite colleges and universities are likely to draw students", which means that, even if aptitude testing becomes commonplace among HE institutions of all kinds, its effects remain particularly pertinent to elite universities.

There have long been concerns about gender differences on performance in aptitude testing in the US (Linn and Hyde, 1989) and, while finding differences in scores by socioeconomic status or gender does not necessarily imply bias (Zwick, 2007, p.20), McDonald et al. (2001b) do identify specific evidence of biases in the SAT, in the US, with "consistent evidence that [it] under-predicts female attainment" once they get to university and more mixed evidence on bias by ethnic groups. Similarly, Wikström and Wikström (2014) present evidence from Sweden that, on average, females perform worse than males in the SweSAT (a national university admissions test), while the opposite is true in measures based on their performance at school. Tannenbaum (2012) argues that one reason for these findings is differing gender styles in test taking, analysing in particular the SAT and differing attitudes to risk.

Although these analyses cannot be extrapolated to the Thinking Skills Assessment, no analysis that I am aware of evaluates whether its predictive power is significantly higher than a baseline of school examination results, nor whether there is evidence of bias in its assessments. The research that has been done specifically into the Thinking Skills Assessment has been restricted to simple analysis of predictive validity with no baseline. Research by Cambridge Assessment (the developers and administrators of the test) sought to examine the extent to which the TSA could predict future academic performance (Emery, 2006; Emery et al., 2006). This was conducted using data from the University of Cambridge courses Computer Science, Economics, Engineering and Natural Science using students who took the TSA in 2003. As is standard practice in evaluating the predictive validity of selection tests, this involved calculating correlations between TSA score and subsequent academic outcomes. In particular, the research finds a correlation between higher marks in the TSA and higher marks in first year university examinations; strong similarities in the candidates that would be rejected by a low TSA cut off score and those rejected under the present selection system; and higher mean TSA scores among those gaining higher degree classification marks in the same examinations.

The authors also state that the correlations, some (but not all) of which are statistically significant, are likely to be an underestimate of the true predictive power since they do not include those who were unsuccessful in getting a place at the university. However, there are potential problems in some of the analysis done because of the data they were able to work with. Rather than having any data where the TSA was administered but not used for selection, the TSA was already in use in the selection process (Emery et al., 2006, p.13). This means that care should be taken in interpretation, especially of the distributions suggesting similarity between those who would be rejected by a TSA cut off and those rejected by the original selection methods.

With rather mixed evidence on predictive validity, we should also consider the wider consequences of introducing an aptitude test. McDonald et al. (2001b, p.53) highlight the importance of this, and draws on the concept of 'consequential validity' (Messick, 1989, p.8). This refers to the wider consequences of introducing the test on other aspects of the admissions process. In this context, we might expect to see a reduced focus on the other information about a candidate that an admissions tutor has: use of aptitude testing may reduce focus on a candidate's examinations results. This might have positive consequences, given known socioeconomic gradients in attainment in such exams. However, that is only the case if the alternative provides a fairer assessment of candidates' ability.

'Consequential validity' also refers to responses to the use of aptitude testing outside the admissions process itself. For example, Wilmouth (1991) argues that students might spend increased time preparing for aptitude tests and less on their academic studies (cited in McDonald et al., 2001b, p.54). This could have a negative knock-on effect on individuals' academic attainment, both in the short term and on their attainment at university. Similarly, Geiser (2008) argues that the education system should reward individuals who work hard throughout their school careers, attaining highly as a result; aptitude testing may incentivise bright individuals to work less hard at achieving high levels of attainment, if they believe they can be successful in gaining access to higher education simply by doing well on a test supposedly designed to assess innate skills.

This paper contributes to the literature by providing evidence on the consequences of aptitude testing for applicants to an elite British university. Given concerns about bias in scores on aptitude tests (Zwick, 2007, p.20) I pay particular attention to these issues, with the paper's research questions as follows:

1. Does use of the TSA have an effect on the proportion of applicants called to interview, the proportion of applicants offered a place, or the proportion of interviewees offered a place?
2. Do these impacts affect high socioeconomic status applicants differently to low socioeconomic status applicants?
3. Do these impacts affect female applicants differently to male applicants?

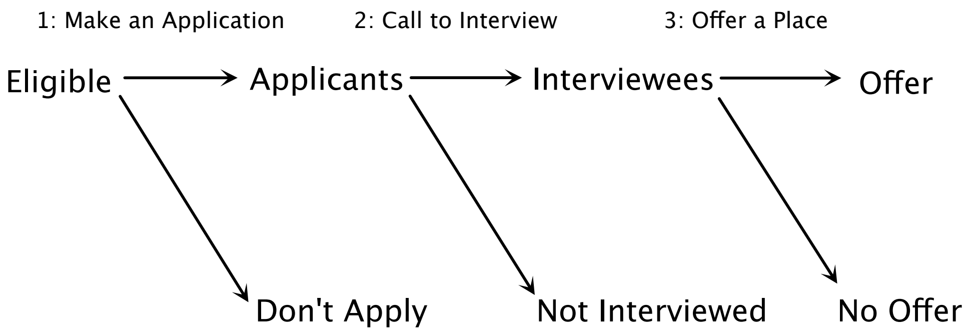
# 3 The admissions process

Unlike at some British universities, the admissions process at the University of Oxford consists of more than one stage, with a shortlist of candidates invited to interview before final admissions decisions are taken. I show the basic form of the admissions process graphically in Figure 1, highlighting three key decision points that make up the process. First, individuals choose whether to apply to Oxford; second, the University chooses which applicants to call to interview[3]; and third, the University chooses whether to offer interviewees a place. Since I am using administrative data from the University (which I will describe further in Section 4), I can analyse the latter two decision points.

Referring back to the idea of 'consequential validity' of using an aptitude test, and the potential for wider societal effects of its introduction, an important part of the story is the impact of the introduction of the TSA on who applies to Economics courses at the University of Oxford.

[3]Starting in 2009, the University has started to introduce use of contextual data in selection to interivew across all subjects. Qualified applicants with various combinations of 'flags' (indicating more challenging circumstances based on prior education and area-based measures) are strongly recommend for interview (University of Oxford, 2014).

Figure 1: Simplified model of the admissions process




Unfortunately, the data available do not allow for the proportion of young people who choose to apply to be modelled since potential applicants are not observed by the university. In any case, the denominator is rather poorly defined. Do we really want to consider the proportion of *all* young people of this age who apply, or restrict attention to subset of 'eligible' applicant? If the latter, who should we regard as an eligible applicant? However, without addressing this matter we might be ignoring significant effects of the policy change. I return to this issue in Section 8.

Thus far, I have described the decision to call candidates to interview, and whether ultimately to offer them a place, as being made by 'the University'. However, to understand who actually makes the decisions it is important to understand the unusual way admissions are organised at the University of Oxford. The University is made up of more than 30 different, fairly autonomous, 'colleges'. Much undergraduate teaching occurs within these colleges, rather than at university level, although students at all colleges, on the same course, study towards the same degree examinations. It is usually one or more of the members of staff who undertake this undergraduate teaching within a college who decide which applicants to invite to interview and, subsequently, which to offer places to. For this purpose, they are referred to as 'admissions tutors'.

A college's admissions tutors' decision over whether to admit an individual is final: University departments cannot overrule college decisions. Most applications for undergraduate courses are made to colleges. However, some individuals do make open applications (which are not to any particular college); these are allocated to a college with a lower applications to places ratio and then proceed on the same basis.

It is worth noting that applicants receiving an offer do not necessarily receive that offer from the college they applied to. The aim of the reallocation process is to ensure that the number of applicants considered by a college is proportional to the number of places available there. Those who are reallocated to other colleges are usually more marginal applicants (since colleges have first refusal on those applicants who apply to them). Roughly 25% of successful applicants are reallocated. The college an individual applies to (or is allocated to if they make an open application) and the college an individual receives an offer from are both recorded in the dataset[4].

All colleges that admit undergraduates admit Economics students[5]. However, the proportion of applicants for Economics and the proportion of offers going to Economics applicants at each college vary greatly (and do not necessarily track one another directly). For example, at the top end, one college received 6.1% of applications to Economics and hosted 8.1% of the university's Economics undergraduates. At the other extreme, one college received just under 1.5% of Economics applications, and went on to host 1% of the university's undergraduate economists.

# 4 Data

I use administrative data from the University of Oxford covering undergraduate admissions made in the years 2005 to 2010. The dataset includes information on all applications to undergraduate courses. This includes applications to Philosophy, Politics and Economics (PPE) and Economics and Management (E&M), the University of Oxford's two main undergraduate degrees in Economics and the subjects for which the aptitude test was introduced; applications to these two courses make up 11% of total applications to Oxford during this period (see Table 1). Throughout the paper I refer to these two courses as Economics, for convenience (although I do explore potentially important differences at various points during the paper).

The progress of applicants through the admissions process is recorded comprehensively in the

[4]I test the robustness of my results to these marginally accepted candidates by redefining receiving an offer to exclude these individuals. In relevant models this does reduce the absolute size of differences and hence statistical significance, but does not materially alter the findings.

[5]I exclude the very small Permanent Private Halls (PPHs), some of which do not offer Economics, and a college that only accepts mature students, since they hence seldom have a school affiliation. Without exclusion these would produce a missing value in proportions of applicants in certain circumstances, resulting in inconsistent sample sizes.

dataset, tracking the individuals who apply, whether they are called to interview, and ultimately whether they are offered a place at the University. Other than details on an applicant's successes or failures (discussed in Section 3), the available data from the process is relatively sparse: it includes their gender, school type (i.e. independent or state), school postcode (which may be linked to data on area level deprivation), and individuals' qualifications, with which to attempt to understand the additional effects attributable to the TSA. Coming from administrative data collected as part of the admissions exercise, the dataset does not include information on the performance of successful individuals once they have been admitted.

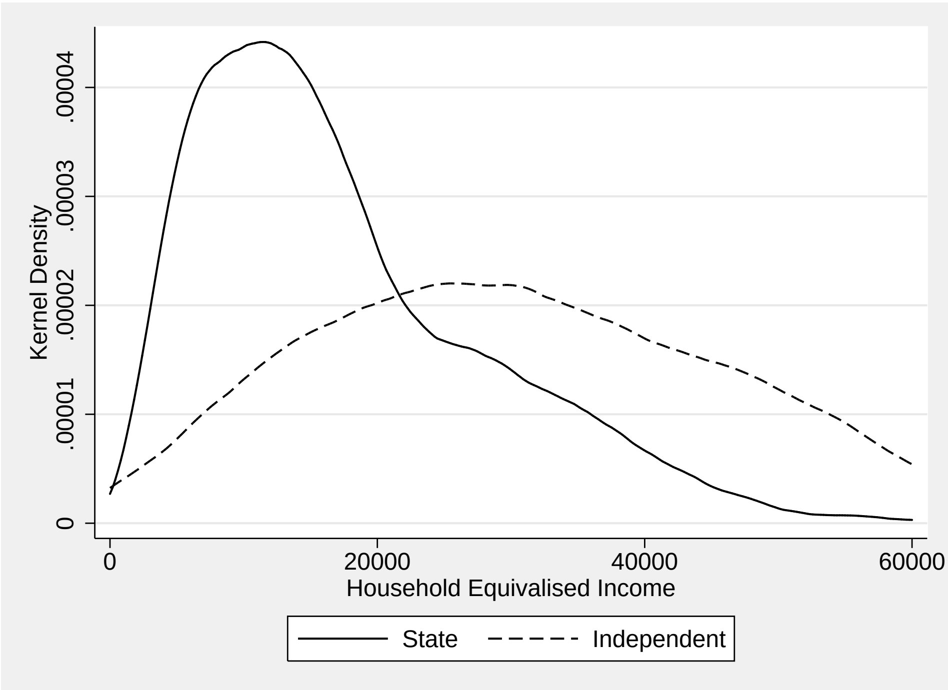
Likewise, as its purpose is to summarise all undergraduate admissions, the dataset does not include information on aspects of the process which are course-specific. Notably, for the purposes of this paper, this means there is no data on individuals' performance in the TSA itself. In any case, this would not, of course, be available for Economics applicants in years prior to its introduction, or for non-Economics applicants in any year. Hence, test scores would not be of use as part of a difference in differences approach. While observable differences in performance in the TSA may explain effects I estimate, the underlying reasons for these are beyond the scope of this paper.

To answer my research questions, I need a proxy for socioeconomic status. Unfortunately, the dataset includes no information on individuals' family backgrounds[6]. I use the variable indicating whether an individual applicant attended an independent school, a state school or neither of these at time of application. I use school type as a proxy for socioeconomic status in this way because of the correlation between the two: in the UK independent schools are primarily fee-paying schools, catering for those from affluent backgrounds. The remainder of the population attends state schools, where funding is provided by the government either through Local Authorities (sometimes referred to as maintained schools) or, increasingly, direct to the schools (which are known as academies). While only about 18% of those in education between the ages of 16 and 18 attend an independent school (Department for Education, 2010), 38% of applicants observed in the dataset are from independent schools.

Attending an independent school does correlate with individuals' socioeconomic status: using

[6]Applications to UK universities are made through the Universities and Colleges Admissions Service (UCAS). As part of this process, individuals are asked to provide information on their ethnic origin, parental education and occupational background. However, these questions are not compulsory. In any event, any responses are not provided to the institutions to which the individual has applied (except in aggregate, and at a later date). As such, they do not form part of this dataset.

Figure 2: Kernel density estimate of the distribution of household equivalised income among young people who apply to university, by whether the young person attends independent school




**Notes:** Own calculations based on data from the Longitudinal Study of Young People in England. Independent school status measured at age 14. Household income equivalised using number of members in household. For more information see Anders (2012b).

data from the Longitudinal Study of Young People in England (LSYPE), I estimate that median household equivalised income for university applicants from state schools is about £14,800, while for those attending an independent school it is just over £31,000[7].

However, there are drawbacks compared to other measures. First, it is a very blunt instrument, providing us with only a binary indicator of status. Second, it proxies socioeconomic status with error: as we can see from Figure 2 there is large overlap in the distributions of household income in households where child is at independent or state school. There will be many reasons for this; for example, in more affluent areas or where schools are selective, more young people from richer backgrounds will attend state schools. Furthermore, in the other direction, individuals from poor backgrounds may attend independent schools, for example supported by bursaries. On the other hand, use of independent school status does have an intuitive appeal. It is both an instantly interpretable distinction and is often the basis for targets regarding fair access that universities negotiate with the UK Government's Office for Fair Access.

The data also include the post codes of the schools that individuals are currently attending (or attended the previous year in the case of applicants who apply shortly after leaving school). By linking with the Income Deprivation Affecting Children Index (IDACI) we may be able to achieve a more nuanced picture of the school's neighbourhood. IDACI "is expressed as the proportion of all children aged 0-15 living in income deprived families" (McLennan et al., 2011, p.22-23). This too will proxy socioeconomic status with error: for example, some schools in deprived neighbourhoods may still attract children from affluent families. However, we can show using other datasets that school IDACI is correlated with an individual's socioeconomic status (see Appendix A).

For the purposes of this analysis I exclude all overseas applicants; those who apply without school affiliation (primarily mature students); and those affiliated to schools where the school type is unavailable for some other reason (about 2% of UK applicants). 63,986 UK applicants for whom we know details about school type remain in the dataset.

Academic attainment of applicants will clearly be an important factor in admissions to any university. In England, the majority of universities use applicants' performance in 'AS Levels',

[7]The LSYPE's measurement of school type is based on a combination of administrative and survey data from approximately age 14. It would be better to measure at age 17 or 18, since a greater proportion of the school population are in independent schools for the two post-compulsory education years leading up to university (about 17.5% vs. 7%). Unfortunately, this is not available: it would make the difference in average income less stark, but would be extremely unlikely to eliminate it.

which are exams taken at around the age of 17, one year into post-compulsory education. In addition, most offers of places will be conditional on applicants achieving a particular set of results in 'A Levels' (these build on AS Levels and are taken two years into post-compulsory education): at the University of Oxford this is typically achieving 3 A-Levels at grade A (the maximum). However, among applicants for courses at Oxford there is very little variability among results in either of these qualifications, with most applicants achieving top grades.

As a result, applicants' performance in General Certificates of Secondary Education (GCSEs) is taken into consideration. In England, these are the predominant examinations taken at the end of compulsory education, usually while individuals are aged 16. In the dataset, I observe the number of GCSEs that applicants have passed and the number of GCSE A*s (the maximum possible grade) that they achieved. As would be expected, GCSE performance differs significantly between applicants, interviewees and those offered a place: the number of GCSE A*s an applicant holds is a good predictor of selection to interview and for an offer[8].

Table 1: Summary statistics of applicants by their school type

| Variable | Overall | Independent | State |
|---|---|---|---|
| Proportion getting an Interview | 0.72 | 0.79 | 0.68 |
| Proportion getting an Offer | 0.26 | 0.30 | 0.23 |
| Proportion of Interviewees getting an Offer | 0.36 | 0.38 | 0.34 |
| Proportion applying to Economics | 0.11 | 0.12 | 0.10 |
| Mean No. of GCSEs passed | 10.28 | 9.99 | 10.46 |
| Mean No. of GCSE A*s | 6.15 | 7.01 | 5.63 |
| N | 63986 | 24470 | 39516 |

*Notes:* Individuals for whom school type is unknown are excluded. Standard errors suppressed as all $\approx 0$.

Table 2: Summary statistics of applicants by their gender

| Variable | Overall | Female | Male |
|---|---|---|---|
| Proportion getting an Interview | 0.72 | 0.72 | 0.72 |
| Proportion getting an Offer | 0.26 | 0.24 | 0.27 |
| Proportion of Interviewees getting an Offer | 0.36 | 0.34 | 0.37 |
| Proportion applying to Economics | 0.11 | 0.07 | 0.14 |
| Mean No. of GCSEs passed | 10.28 | 10.29 | 10.28 |
| Mean No. of GCSE A*s | 6.15 | 6.48 | 5.85 |
| N | 63986 | 30985 | 33001 |

*Notes:* Individuals for whom school type is unknown are excluded. Standard errors suppressed as all $\approx 0$.

Applicants from independent schools have different observable characteristics, on average.

[8]Using a simple linear probability model containing only the number of GCSE A*s held by a candidate as a continuous regressor, I estimate that each additional GCSE A* increases a candidate's probability of being offered a place by approximately 4.6 percentage points. The t-statistic on this coefficient is 83.3 and the overall model has an $R^2$ of 0.10. I get very similar results with a linear probability model of selection to interview.

For example, in Table 1 we can see that they receive on average fewer GCSEs. While this may seem counter-intuitive, independent schools may encourage their pupils to take slightly fewer GCSEs to maximise performance on those they do take. Indeed, applicants from independent schools have more GCSEs awarded A*s (the highest grade). In addition, a larger proportion of independent school applicants apply to Economics than do state school applicants. Likewise, there are observable differences, on average, between male and female applicants. Female applicants are just as likely to get an interview, but less likely to receive an offer. This is despite having a statistically significantly higher mean number of GCSEs awarded A*s than their male counterparts. They are also half as likely to apply to Economics as male applicants.

Table 3: Summary statistics of applicants by subject group applied to

| Variable | Overall | Economics | Others |
|---|---|---|---|
| Proportion getting an Interview | 0.72 | 0.69 | 0.72 |
| | ( 0.00) | ( 0.01) | ( 0.00) |
| Proportion getting an Offer | 0.26 | 0.22 | 0.26 |
| | ( 0.00) | ( 0.00) | ( 0.00) |
| Proportion of Interviewees getting an Offer | 0.36 | 0.31 | 0.36 |
| | ( 0.00) | ( 0.01) | ( 0.00) |
| Proportion from Independent school | 0.38 | 0.44 | 0.38 |
| | ( 0.00) | ( 0.01) | ( 0.00) |
| Proportion who are female | 0.48 | 0.33 | 0.50 |
| | ( 0.00) | ( 0.01) | ( 0.00) |
| Mean No. of GCSEs passed | 10.28 | 10.26 | 10.29 |
| | ( 0.01) | ( 0.02) | ( 0.01) |
| Mean No. of GCSE A*s | 6.15 | 6.33 | 6.13 |
| | ( 0.01) | ( 0.04) | ( 0.01) |
| N | 63986 | 6904 | 57082 |

*Notes:* Individuals for whom school type is unknown are excluded. Standard errors in parentheses.

Less obviously, admissions statistics and average attainment of applicants also differ significantly by course choice. Table 3 shows summary statistics for the two groups, Economics and all other subjects. It shows us that Economics applicants are already less likely to get an interview than other subjects, and are less likely ultimately to receive an offer (these differences are statistically significant). We should note that since the supply of places is effectively fixed: as the proportion getting an offer is driven by differences in demand there is no particular reason to expect the proportions to be the same across courses. In addition, there is a larger proportion of applicants from independent schools for Economics. Importantly for this work, applicants for Economics have, on average, statistically significantly fewer GCSE A*s than applicants for other subjects; conversely, those who receive offers for Economics courses have more GCSE A*s than those offered a place for other subjects, again on average. This suggests GCSE performance is a particularly important predictor for Economics, relative to other subjects: I attempt to mitigate this problem by controlling for GCSE performance using least squares regression as part of my analysis.

Given their importance in the admissions process, it is also important to consider the differences between colleges. Within the University of Oxford, colleges have differing academic reputations. It seems plausible that this may affect the quality of applicants to, and selectivity of, individual colleges. The University-produced ‘Norrington score’ may capture some of this. According to the University website it “provides a way of measuring the performance of students at each college in the end of university exams” (University of Oxford, 2013). The Norrington score is based on the classifications of undergraduate degrees awarded, attaching a score of 5 to a first class degree, 3 to an upper second class degree, 2 to a lower second class degree, 1 to a third class degree and 0 to a pass. It is calculated by dividing the total college score by the total possible score the college could attain and multiplying by 100 to yield a percentage. I assign each college’s Norrington score to the group of applicants in the autumn following the examinations on which the score is based. This means that it will be the most recent piece of information on college quality that applicants and interviewers will have.

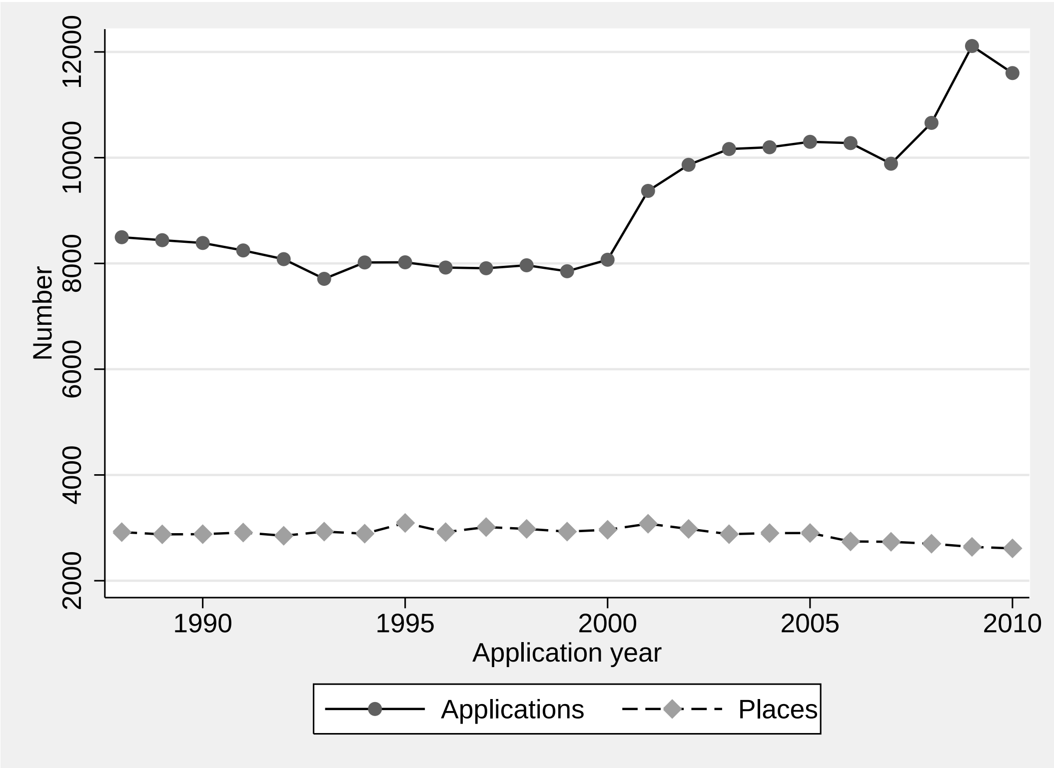
# 5 Trends in admissions and introduction of the TSA

The University of Oxford has experienced a large increase in applications for all courses since roughly the year 2000, as can be seen in Figure 3. After roughly 10 years of receiving approximately 8,000 applications from UK students each year, this grew rapidly by about 50% to a peak of around 12,000 in 2009, although it fell back somewhat in 2010. This has been driven particularly by a large increase in the number of applications from state school pupils during this period (see Figure 4), rising from under 4,500 to about 7,500. However, there has been no corresponding increase in the number of offers made, which have continued at around 3,000 and, if anything, declined slightly as more offers have gone to overseas applicants. It follows that getting a place has become considerably more competitive.

Over the shorter period for which I can observe subject-specific figures[9], Economics is no ex-

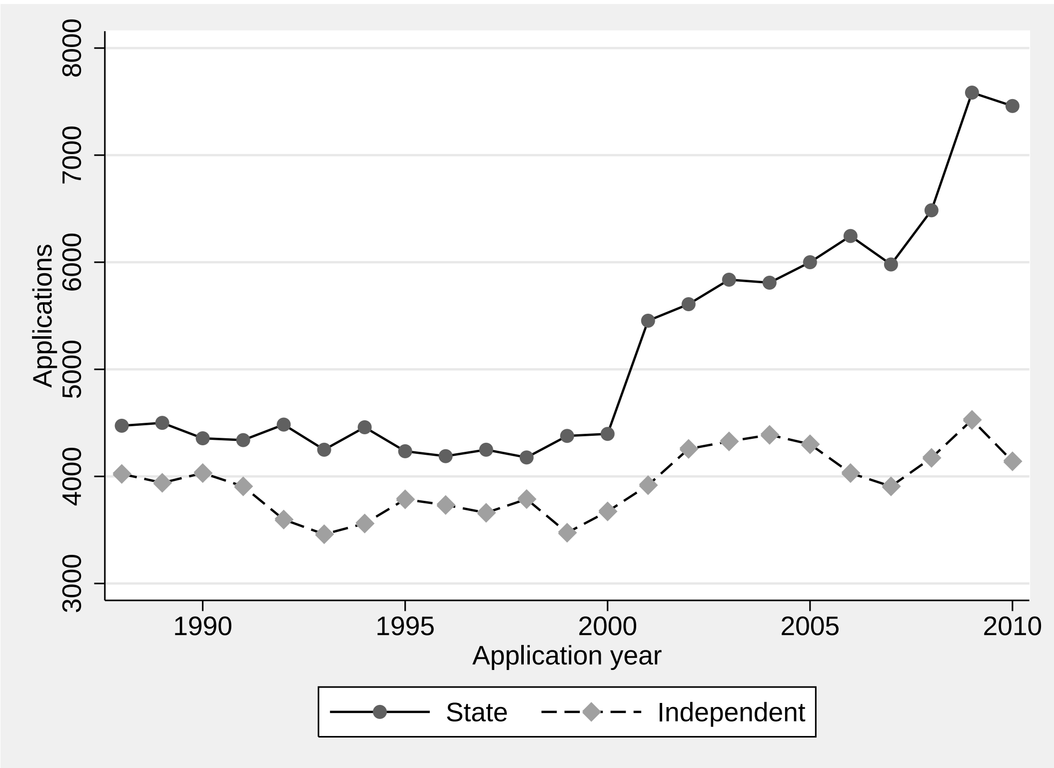
[9]It should be noted that this covers only about half the period of the large rise in applications to the University in general.

Figure 3: Number of applications from and offers given to UK students, by year



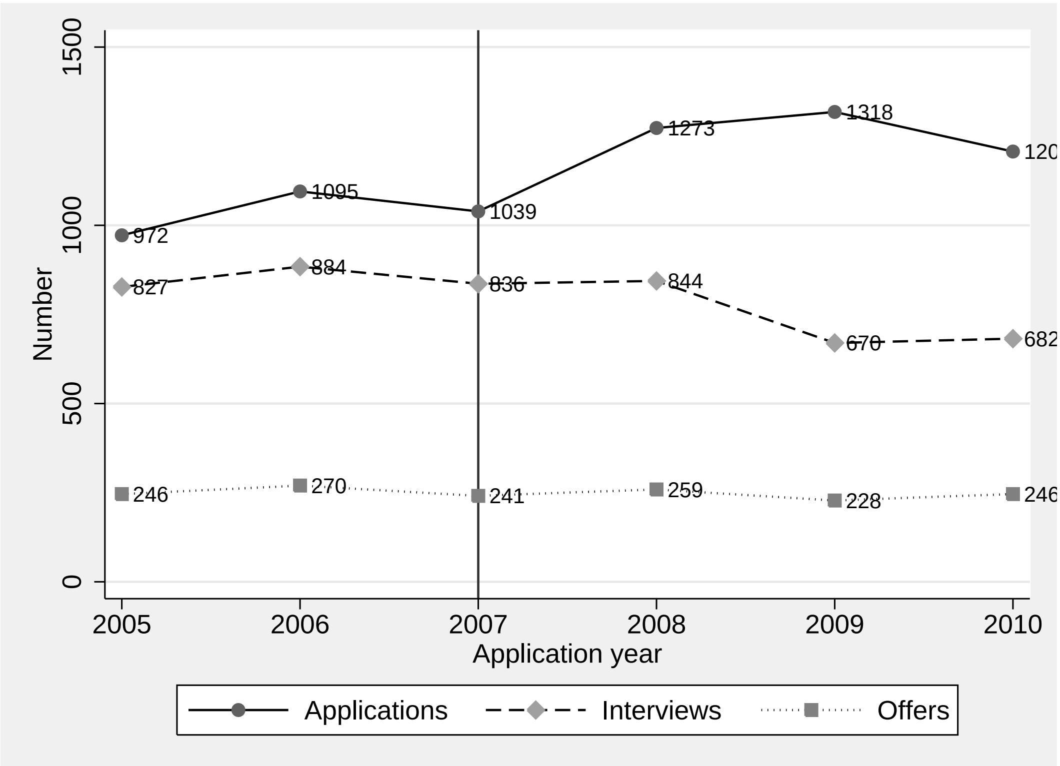

*Notes:* Source: Oxford University Admissions Statistics, across all subjects. Individuals for whom school type is unknown are excluded.

Figure 4: Number of applications from UK students, by year and school type




*Notes:* Source: Oxford University Admissions Statistics, across all subjects. Individuals for whom school type is unknown are excluded.

Figure 5: Number of applications to, interviews for and offers for Economics from UK students, by year




*Notes:* Sample size: 6,904. Individuals for whom school type is unknown are excluded. Vertical line indicates the year 2007, when test was administered but not used to inform decisions. In years before the line the test was not used; and in years after the test was used as part of the admissions process.

ception to the pattern of increasing applications. Figure 5 shows that the number of applications has risen from 972 in 2005 to a peak of 1,318 in 2009 (with a similar slight reduction in 2010 as that seen in the overall figures, but still above that seen between 2005-2007). Again, the number of places to study Economics awarded to UK students has not risen alongside this.

Faced with this large increase in the number of applications, and the labour-intensive nature of the interview stage of the admissions process, the decision was taken to introduce a guideline for the number of interviews a college should conduct per place it had available. In Figure 5, we observe this fall in the number of interviews from 836 in 2007 to 682 in 2010. This is a sizeable difference; with potential knock-on effects. The TSA was introduced at the same time in order to support this policy, providing admissions tutors with additional information with which to select applicants to call to interview. As such, the test was a requirement for all individuals applying to these subjects; this is unlike some institutions' use of the TSA, where it is administered only to interviewees (Admissions Testing Service, 2013a). Candidates sit the TSA at their school[10] on a date in early November, just under a month after the deadline for applications. Results are available to admissions tutors shortly afterwards, but are not released to the candidates until early the following year (after interviews have been conducted and offers made).

The TSA was introduced in a phased approach. Applicants to Philosophy, Politics and Economics (PPE) at the University first sat the TSA in 2007. A complication in 2007 is that the test was administered to PPE applicants, but the results were not released to admissions tutors until after they had selected which applicants to call for interview. As such, it was not used to make decisions on who to call to interview, but was available to make decisions on which applicants to offer places to. This means we might expect to see some of the effects of the policy (for example due to changing behaviour by applicants), but not others (due to changing behaviour by admissions tutors in selecting candidates for interview). Applicants to Economics and Management (E&M) first sat the test in 2008. Unlike in PPE, the results of the TSA were available to admissions tutors when deciding which applicants to call for interview from that first year. However, in a different complication the guideline for the number of interviews per place was not introduced for TSA until 2009. These differences in implementation have the

[10]If the school is not willing to administer the test then candidates may take it at an approved test centre, usually another school or college nearby.

potential to distort the analysis. Since the impact of the test is our fundamental interest, I elect to exclude 2007 from the analysis. Since applicants do sit the test in 2008 and the results are available throughout the process to admissions tutors, I do not exclude it. However, the later implementation of the target number of interviews per place in E&M means there was a relatively larger number of E&M than PPE interviews in 2008: as such E&M interviews will weigh particularly heavily in that year. I am careful to discuss explore and discuss potential implications for the results in 2008[11].

In my analysis, I exploit the fact that in the data there are two years where the aptitude test was not administered (2005 and 2006); and three years where it was administered to all Economics applicants (2008, 2009 and 2010). The policy has then continued in more recent years, but I do not have access to the data from this period. This natural experiment presents an opportunity to evaluate the effects stemming from this policy change, with no other major confounding policy changes affecting admissions having been undertaken at this time, to my knowledge[12].

As noted above, since 2000 there have been large increases in the number of applications to the University, but no increase in the number of offers made. Estimating the impact of the TSA just by looking at characteristics before the change in policy and comparing them to the same characteristics afterwards would be biased by the general downward trend in the proportion of applicants receiving an offer. Instead, I estimate the impact using a difference in differences (DiD) framework. This attempts to control for any general trends using the trends seen in subjects where the TSA was not introduced, hence attempting to isolate the changes in our outcome measures of interest that are due to the introduction of the TSA. The identifying assumption is that changes in the outcome variables for Economics applicants, over and above those seen among applicants to other subjects, are due to the introduction of the TSA: this requires that the trends in the treatment and control groups are the same, the so-called 'common trends' assumption. For most of my analysis, the 'treatment' group is Economics and the 'control' group are all other subjects. The policy of interest, the introduction of the TSA, is

[11]Although not reported in this paper, I do also run models including 2007 to check for unexpected effects, and run models that estimate the effect for PPE and E&M application processes separately. These do not alter the main thrust of the findings.

[12]Undergraduate tuition fees rose from £1000 to a maximum of £3000 in the academic year 2006/7. The majority of applications for that year's entry would be made in 2005, at the very beginning of this dataset. As such, any changes in application behaviour associated with this policy change should not confound the analysis in this paper, although they could affect pre-treatment trends.

'off' in 2005 and 2006, and 'on' in 2008, 2009 and 2010.

Common trends are more likely if the 'control' group (other subjects) has similar observable characteristics to the Economics 'treatment' group. In Section 4, I discussed some of the differences between the profile of the average Economics applicant and the average applicant to other subjects, noting in particular differences in the average academic attainment between the two groups. However, the subject groups are not so different that it casts doubt on the validity of other subjects as a 'control' group. I also use a more restricted control group as a robustness check, which I discuss further in Section 9.

Table 4: Proportion of applicants who receive an offer, proportion of applicants who receive an interview, and proportion of interviewees who receive an offer, by year and subject group: difference in differences estimates

| **Apply $\rightarrow$ Offer** | **Policy Off** | **Policy On** | **Difference** |
|---|---|---|---|
| Economics | 0.250 | 0.193 | -0.057 |
| | ( 0.013) | ( 0.010) | ( 0.012)*** |
| Others | 0.284 | 0.241 | -0.043 |
| | ( 0.006) | ( 0.006) | ( 0.005)*** |
| Difference | -0.034 | -0.048 | -0.014 |
| | ( 0.014)*** | ( 0.011)*** | ( 0.013) |
| **Apply $\rightarrow$ Interview** | **Policy Off** | **Policy On** | **Difference** |
| Economics | 0.828 | 0.578 | -0.250 |
| | ( 0.015) | ( 0.016) | ( 0.023)*** |
| Others | 0.788 | 0.677 | -0.111 |
| | ( 0.007) | ( 0.007) | ( 0.006)*** |
| Difference | 0.040 | -0.099 | -0.139 |
| | ( 0.016)*** | ( 0.017)*** | ( 0.024)*** |
| **Interview $\rightarrow$ Offer** | **Policy Off** | **Policy On** | **Difference** |
| Economics | 0.302 | 0.334 | 0.032 |
| | ( 0.016) | ( 0.012) | ( 0.017)* |
| Others | 0.361 | 0.356 | -0.004 |
| | ( 0.006) | ( 0.006) | ( 0.005) |
| Difference | -0.059 | -0.023 | 0.036 |
| | ( 0.017)*** | ( 0.013)* | ( 0.018)** |

*Notes:* Analysis excludes individuals for whom school type is unknown. Policy Off in 2005, 2006 and 2007; Policy On in 2008, 2009 and 2010. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate statistical significance: $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$. Sample sizes: Apply $\rightarrow$ Offer: 63986 Apply $\rightarrow$ Interview: 63986 Interview $\rightarrow$ Offer: 46106

Table 4 shows the change in the proportion of applicants getting interviews and places from before to after the policy change, for Economics and other subjects. While there is a significant reduction in the proportion of Economics applicants receiving offers, this is matched by a similar fall in the proportion getting an offer in other subjects.

By contrast, the reduction in proportion of Economics applicants getting an interview is significantly larger than that seen in other subjects, driven by the introduction of a guideline number of interviews per available place. Table 4 shows a simple estimate of the effect of the policy's introduction on the proportion of applicants who receive an interview: a 11.5 percentage point reduction. When coupled with no effect on the overall proportion receiving an offer, this implies that the policy must have resulted in an increase in the proportion of interviewees getting an offer. This is indeed what we see, with the proportion of Economics interviewees receiving an offer increasing, even as this statistic falls for other subjects. Our simple estimate of the policy impact is a 5.4 percentage point increase in the proportion of interviewees who receive an offer.

A reduction in the proportion of applicants who are called to interview would appear to be an increase in efficiency of the admissions process. However, it could be that this is a trade-off against other aims: selecting the highest quality applicants for the course and doing so without bias from applicants' other characteristics. Testing the first of these might be possible, but would require data on candidates' performance in their final examinations, which is not available in the dataset. However, I now shed some light on the second aim.

The large reduction in the proportion of applicants called for interviews clearly allows for the possibility of relative changes in the proportion of applicants from different genders or school types who are called for interview. Neither do the findings so far rule out the possibility of the policy having an effect on the proportion of applicants receiving an offer and coming from a particular group, since countervailing effects could offset one another.

To consider these matters, I present versions of Table 4 that separate out the overall effect of the policy into separate effects by our groups of interest. For the exposition of this analysis, I concentrate on effects by school type. However, it is easy to see how this is translated to analyse differences by gender.

For these purposes, instead of using the overall proportion of applicants who get a place, I analyse two sets of proportions: one where the numerator consists of only those getting an offer (or an interview) <u>and</u> coming from an independent school; and the other where the numerator consists of only those getting an offer (or an interview) <u>and</u> coming from a state school (on the right side of the table). In both cases, the denominator remains, as for Table 4, all applicants (or interviewees, in the case of Offer | Interview).

To make this clearer, I define the following notation:

$A_I =$ Number of applicants from independent schools

$A_S =$ Number of applicants from state schools

$I_I =$ Number of interviewees from independent schools

$I_S =$ Number of interviewees from state schools

$O_I =$ Number of offers to individuals from independent schools

$O_S =$ Number of offers to individuals from state schools

The proportions reported in the table are as follows:

**Proportion of applicants receiving an offer** : Independent: $\frac{O_I}{A_I+A_S}$ State: $\frac{O_S}{A_I+A_S}$

**Proportion of applicants receiving an interview** : Independent: $\frac{I_I}{A_I+A_S}$ State: $\frac{I_S}{A_I+A_S}$

**Proportion of interviewees receiving an offer** : Independent: $\frac{I_I}{I_I+I_S}$ State: $\frac{O_S}{I_I+I_S}$

This DiD analysis is presented in Table 5. How do these proportions relate to the previous analysis and to one another? The proportions reported in Table 4 were of the form $\frac{I_I+I_S}{A_I+A_S}$ (this particular example is the proportion of applicants called to interview). The proportions separated by school type are a simple decomposition of this overall proportion, since $\frac{I_I}{A_I+A_S} + \frac{I_S}{A_I+A_S} = \frac{I_I+I_S}{A_I+A_S}$. Ensuring that the outcome variables for the independent and state school analyses have the same denominator means we can easily compare the DiD estimates from each to see whether there are differential effects of the policy on applicants from the two school types.

In the case of the overall proportion receiving an offer, the story does not immediately seem more complex than suggested by Table 4. In the top panel, we do not see a statistically significant change in the proportion of all applicants who are successful and come from either school type.

However, looking at the middle panel, at first look there would appear to be a difference between the effects on the proportion of all applicants called to interview by school type. The difference in difference estimate of the effect on the proportion relating to state school interviewees is a reduction of 5.4 percentage points, while the relevant effect relating to those

Table 5: Proportion of all applicants who receive an offer, proportion of all applicants who receive an interview, and proportion of all interviewees who receive an offer, by school type, year and subject group: difference in differences estimates

| | Independent | | | State | | |
|---|---|---|---|---|---|---|
| **Apply $\rightarrow$ Offer** | **Policy Off** | **Policy On** | **Difference** | **Policy Off** | **Policy On** | **Difference** |
| Economics | 0.123 | 0.091 | -0.032 | 0.127 | 0.102 | -0.025 |
| | ( 0.010) | ( 0.006) | ( 0.009)*** | ( 0.010) | ( 0.008) | ( 0.008)*** |
| Others | 0.128 | 0.106 | -0.022 | 0.156 | 0.135 | -0.020 |
| | ( 0.006) | ( 0.004) | ( 0.004)*** | ( 0.006) | ( 0.004) | ( 0.004)*** |
| Difference | -0.005 | -0.015 | -0.009 | -0.029 | -0.034 | -0.005 |
| | ( 0.012) | ( 0.007)** | ( 0.010) | ( 0.012)*** | ( 0.009)*** | ( 0.009) |
| | Independent | | | State | | |
| **Apply $\rightarrow$ Interview** | **Policy Off** | **Policy On** | **Difference** | **Policy Off** | **Policy On** | **Difference** |
| Economics | 0.392 | 0.268 | -0.124 | 0.436 | 0.310 | -0.126 |
| | ( 0.022) | ( 0.015) | ( 0.014)*** | ( 0.020) | ( 0.020) | ( 0.017)*** |
| Others | 0.321 | 0.283 | -0.038 | 0.466 | 0.394 | -0.072 |
| | ( 0.014) | ( 0.010) | ( 0.008)*** | ( 0.014) | ( 0.010) | ( 0.009)*** |
| Difference | 0.071 | -0.015 | -0.085 | -0.030 | -0.084 | -0.054 |
| | ( 0.026)*** | ( 0.017) | ( 0.016)*** | ( 0.024) | ( 0.022)*** | ( 0.019)*** |
| | Independent | | | State | | |
| **Interview $\rightarrow$ Offer** | **Policy Off** | **Policy On** | **Difference** | **Policy Off** | **Policy On** | **Difference** |
| Economics | 0.148 | 0.158 | 0.010 | 0.153 | 0.176 | 0.023 |
| | ( 0.012) | ( 0.010) | ( 0.013) | ( 0.013) | ( 0.011) | ( 0.010)* |
| Others | 0.163 | 0.156 | -0.007 | 0.198 | 0.200 | 0.002 |
| | ( 0.007) | ( 0.006) | ( 0.005) | ( 0.007) | ( 0.006) | ( 0.005) |
| Difference | -0.015 | 0.002 | 0.016 | -0.044 | -0.024 | 0.020 |
| | ( 0.014) | ( 0.011) | ( 0.014) | ( 0.015)*** | ( 0.012)** | ( 0.011)* |

*Notes:* Outcome variables reported are (**Apply $\rightarrow$ Offer**) proportion of all applicants who receive an offer and come from given school type, (**Apply $\rightarrow$ Interview**) proportion of all applicants who receive an interview and come from given school type, and (**Interview $\rightarrow$ Offer**) proportion of all interviewees who receive an offer and come from given school type. Analysis excludes individuals for whom school type is unknown. Policy Off in 2005, 2006 and 2007; Policy On in 2008, 2009 and 2010. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate statistical significance: $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$. Sample sizes: Apply $\rightarrow$ Offer: 63986; Apply $\rightarrow$ Interview: 63986; Interview $\rightarrow$ Offer: 46106.

from independent schools is a reduction of 8.5 percentage points. There are reductions in both these proportions, but the effect on the proportion of all interviewees being called to interview and coming from independent school is larger; the estimated effect is roughly 3 percentage points greater in magnitude. Nevertheless, we cannot reject the null hypothesis of no difference between these two estimates at the conventional 5% level (although we can at the 10% level).

Finally, turning to the bottom panel of Table 5 the proportion of interviewees who receive offers and come from state schools is estimated to increase slightly more than the proportion of all interviewees who are successful and come from independent schools (2.0 percentage points, compared with 1.6 percentage points). However, a simple t-test confirms that the estimated effects are not significantly different from one another.

In Table 6 I report the same analysis split by gender. I do not find statistically significant differences in the overall effect of introducing the TSA on the proportion of applicants getting an offer by gender, although if there is any difference it is to the detriment of female applicants. However, we do again see differences in the results by gender when considering the two separate stages of the admissions process. Considering first the proportion of applicants offered an interview, we see that the proportion of all applicants offered an interview and who are female has declined by 5.5 percentage points, compared to a larger decline of 8.6 percentage points in the proportion of all applicants offered an interview and who are male. However, we cannot reject the null hypothesis of no difference between these two estimates at the conventional 5% level (although we can at the 10% level).

In any case, the difference appears to be offset at the latter stage of the admissions process. We saw above that the proportion of interviewees getting an offer increased in response to the introduction of the TSA (offsetting the falling numbers getting an interview): the results by gender suggest that this is entirely driven by the proportion of all interviewees receiving an offer and who are men (4.4 percentage point increase, compared to a very small decrease for females). This difference does appear to be statistically significant at the 5% level. Given that the aptitude test is primarily used to select candidates for interview, finding an effect at the latter stage of the admissions process may seem unexpected. However, an indirect effect of this type is possible. One explanation is that the TSA is filtering out the kind of female interviewees who previously went on to perform well at interview and hence receive an offer.

Table 6: Proportion of all applicants who receive an offer, proportion of all applicants who receive an interview, and proportion of all interviewees who receive an offer, by gender, year and subject group: difference in differences estimates

| | Female | | | Male | | |
|---|---|---|---|---|---|---|
| **Apply → Offer** | **Policy Off** | **Policy On** | **Difference** | **Policy Off** | **Policy On** | **Difference** |
| Economics | 0.089 | 0.057 | -0.032 | 0.161 | 0.136 | -0.025 |
| | ( 0.006) | ( 0.005) | ( 0.007)*** | ( 0.010) | ( 0.008) | ( 0.011)*** |
| Others | 0.135 | 0.115 | -0.020 | 0.149 | 0.126 | -0.023 |
| | ( 0.003) | ( 0.003) | ( 0.003)*** | ( 0.005) | ( 0.005) | ( 0.005)*** |
| Difference | -0.047 | -0.059 | -0.012 | 0.012 | 0.010 | -0.002 |
| | ( 0.007)*** | ( 0.006)*** | ( 0.007) | ( 0.011) | ( 0.009) | ( 0.012) |
| | Female | | | Male | | |
| **Apply → Interview** | **Policy Off** | **Policy On** | **Difference** | **Policy Off** | **Policy On** | **Difference** |
| Economics | 0.269 | 0.167 | -0.102 | 0.558 | 0.411 | -0.147 |
| | ( 0.012) | ( 0.008) | ( 0.013)*** | ( 0.016) | ( 0.012) | ( 0.022)*** |
| Others | 0.391 | 0.342 | -0.049 | 0.396 | 0.335 | -0.062 |
| | ( 0.008) | ( 0.007) | ( 0.008)*** | ( 0.010) | ( 0.007) | ( 0.009)*** |
| Difference | -0.122 | -0.175 | -0.053 | 0.162 | 0.076 | -0.086 |
| | ( 0.014)*** | ( 0.011)*** | ( 0.015)*** | ( 0.019)*** | ( 0.014)*** | ( 0.024)*** |
| | Female | | | Male | | |
| **Interview → Offer** | **Policy Off** | **Policy On** | **Difference** | **Policy Off** | **Policy On** | **Difference** |
| Economics | 0.107 | 0.098 | -0.009 | 0.195 | 0.236 | 0.041 |
| | ( 0.008) | ( 0.007) | ( 0.010) | ( 0.012) | ( 0.010) | ( 0.015)*** |
| Others | 0.172 | 0.170 | -0.001 | 0.189 | 0.186 | -0.003 |
| | ( 0.004) | ( 0.004) | ( 0.004) | ( 0.006) | ( 0.006) | ( 0.006) |
| Difference | -0.065 | -0.072 | -0.008 | 0.006 | 0.050 | 0.044 |
| | ( 0.009)*** | ( 0.008)*** | ( 0.011) | ( 0.013) | ( 0.012)*** | ( 0.016)*** |

*Notes:* Outcome variables reported are (**Apply** → **Offer**) proportion of all applicants who receive an offer and come from given school type, (**Apply** → **Interview**) proportion of all applicants who receive an interview and come from given school type, and (**Interview** → **Offer**) proportion of all interviewees who receive an offer and come from given school type. Analysis excludes individuals for whom school type is unknown. Policy Off in 2005, 2006 and 2007; Policy On in 2008, 2009 and 2010. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate statistical significance: $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$. Sample sizes: Apply → Interview: 63986; Apply → Offer: 63986; Interview → Offer: 46106.

I investigate such explanations further while discussing the results from the regression models in Section 7.

So far, these results answer my research questions in the following ways: they do not suggest an impact on the proportion of applicants offered a place, but do reflect the negative impact on the proportion of applicants called to interview caused by the introduction of a target number of interviews per place. As such, we also see an offsetting increase in the proportion of interviewees offered a place. I find some limited evidence of differences in these impacts by the socioeconomic status of applicants, with the proportion of applicants getting an interview and coming from an independent school declining more than for its state school counterpart. In addition, there is evidence of differential effects on the proportion of applicants getting an interview and the proportion of interviewees getting an offer by gender. Nevertheless, these results should not overshadow the finding that in neither of these cases (differences by school type or gender) do we find a statistically significant overall difference in the proportion of all applicants who receive an offer.

However, this simple analysis has limitations, which I aim to check and/or relax, as appropriate, using regression analysis below.

# 6 Regression analysis

DiD estimates may be conveniently recovered using least squares regression. In addition, regression analysis allows increased model flexibility compared to those I have used thus far. I use this flexibility to check for different effects by year and to control for college-, course- and time-varying covariates that could affect the validity of the common trends assumption.

As discussed in Section 1, decisions about who to admit are made by admissions tutors at each college. Given their importance, I perform regression analysis using colleges as the unit of analysis. I collapse individual applicant records into college-level averages, also maintaining separate observations by year and course group. After exclusions, the data include 29 colleges, six years and two course groups (Economics and Others). This gives 348 college, year, course group combinations forming available observations for the regression analysis. In all specifications, year variables are grouped in some way, reducing the number of observations to those shown in later results tables.

I weight the observations to take account of the average number of applicants a college receives in a year across the whole period from 2005 to 2010. Colleges vary significantly in size so, as the underlying research questions are about the effects on applicants, weighting to be representative of the numbers of applicants is appropriate. Failure to do this would implicitly give each college an equal weighting, potentially vastly over-exaggerating the influence of small colleges on the overall results. Analytic weights are used, as these take into account the fact that the observations are means, made up of observations of individuals' characteristics and progress through the admissions process[13].

I begin by replicating the analysis in Section 5 above in a regression framework, using an equation of the form shown in Equation 1. As a result of the weighting strategy, we would not expect the point estimates to be quite identical to those in earlier analysis.

$$
\begin{aligned}
Y_{jt} = \alpha &+ \beta_p \text{Treated}_j \\
&+ \gamma \text{Policy On}_t \\
&+ \delta \text{Treated}_j * \text{Policy On}_t + \varepsilon_{jt}
\end{aligned}
\tag{1}
$$

where $Y_{jt}$ is the outcome of interest at college $j$ in year $t$; Treated are dummy variable indicating the two treatment groups (both PPE and E&M); Policy On is a dummy variable set to 0 in years 2005 and 2006, and 1 in 2008, 2009 and 2010; and $\varepsilon$ is an error term (which I discuss further below).

The coefficients on Treated ($\beta$) control for pre-existing differences between applicants to these and other subjects; the coefficient on Policy On ($\gamma$) controls for general trends in the variables relative to the base years of 2005 and 2006; and the coefficient on the interaction term between the Treated and Policy On variables ($\delta$) allows us to recover the impact of the TSA, under the identifying assumption of common trends.

However, regression analysis makes it easy to introduce more flexibility than I have allowed for so far; I take advantage of this in various ways. First, I allow for different effects each year by replacing the Policy On dummy variables with a set of year dummies. Equation 2 shows the

[13] This echoes the approach by Card (1992), who estimates the impact of minimum wages using observations from 51 states, weighting these by the average size of the sample for relevant workers in each state.

form of equation used.

$$
\begin{aligned}
Y_{jt} &= \alpha + \beta \text{Treated}_j \\
&\quad + \gamma_8 2008_t + \gamma_9 2009_t + \gamma_{10} 2010_t \\
&\quad + \delta_8 \text{Treated}_j * 2008_t + \delta_9 \text{Treated}_j * 2009_t + \delta_{10} \text{Treated}_j * 2010_t + \varepsilon_{jt} \qquad (2)
\end{aligned}
$$

where 2008, 2009 and 2010 are dummy variables indicating cohorts where the policy is on.

The interpretation for Equation 2 is very similar to that for Equation 1. The coefficient on Treated ($\beta$) still controls for pre-existing differences between applicants to Economics and other subjects; the coefficients on 2008, 2009 and 2010 ($\gamma$) control for general trends in the variables relative to the base years of 2005 and 2006; and the coefficients on the interaction terms between the the Treated (just a combination of PPE and E&M variable) and year variables ($\delta_8$, $\delta_9$ and $\delta_{10}$) allow us to recover the estimated impact of the TSA for each of these treatment years.

I also use regression to include additional college-, course-, and time-varying covariates. Including these covariates aims to help control for omitted college- and course-specific trends in the outcome variables that could otherwise undermine the common trends assumption. Firstly, I include measures of the average academic performance of applicants from our groups of interest (applicants from independent and state schools for school type analysis; male and female applicants for analysis by gender) to each course group at each college (using the number of GCSEs and the number of GCSE A*s held by the mean applicant from each school type). These aim to control for changes in the success of candidates from each school type that are due to observable differences in their prior academic attainment. Secondly, I include an annual measure of the performance of the college's undergraduates at the end of their degrees (using the Norrington score, discussed in Section 4). This aims to control for the possibility that the quality of applicants to a college is affected by its academic reputation. I use a regression equation very similar to that in Equation 2, except for the addition of this vector of college-level controls.

As is common in DiD analysis, various aspects of the data are problematic for classical statistical inference (Bertrand et al., 2004). However, there is a growing literature on inference in such circumstances (Brewer et al., 2013). In particular, I adapt advice from Angrist and

Pischke (2009, ch. 8) in my approach to obtaining appropriate standard errors. First, while admissions tutors are college- and subject-specific, some courses have more than one subject area. It follows that there may be cases where the same admissions tutor makes decisions in different courses. As such, I allow for clustering between courses, other than between the treatment and control groups (i.e. Economics-related subjects and others). Given that most courses do have different admissions tutors, this is a very conservative approach[14]. Second, repeated observations across several years, often likely with the same admissions tutor with persistent preferences over time, makes autocorrelation/serial correlation likely (Kennedy, 2008, p.118).

As the observations are in the form of college, year, course group combinations, this already allows for clustering within college and course group combinations. However, it assumes independence by year. As such, I use Stata's cluster option to define clusters as the 58 college and course group combinations, allowing for serial correlation.

# 7 Results

Given this paper's particular focus on the potential for differential effects on applicants by their socioeconomic background or gender, I take as given the picture of the reduction in proportion of applicants who are called for interview and offsetting increase in the proportion of interviewees who are offered a place[15]. I proceed immediately to analyse whether evidence exists of differential effects for applicants, beginning with school type before turning to gender.

Results are presented in tables for each stage of the admissions process, with regression models in numbered columns. In each column, the DiD estimates of policy impact are shown either by rows giving the interaction between Economics and policy on ($\delta$) or by rows giving the interaction between Economics and treatment years ($\delta_8$, $\delta_9$ and $\delta_{10}$) depending on the model. I then report the differences between the DiD estimated effects for each pair of models, with the

[14]Nevertheless, we might wish to allow clustering even between Economics and other subjects. However, in doing so we reduce the number of clusters to equal the number of colleges (after the exclusions described above): this is only 29 clusters. This is short of the minimum of 42 recommended for standard clustering techniques by Angrist and Pischke (2009). The 'wild bootstrap t-procedure' (Cameron et al., 2008) is more effective at avoiding type II errors with such a small number of clusters. Performing inference even on this extremely conservative basis does not materially alter the statistical significance of my results. I implement this using the command by Bansi Malde, available from http://www.ifs.org.uk/publications/6231

[15]I do estimate these regression models to check the robustness of the analysis in Table 4, but do not report the results in this paper.

statistical significance of the differences indicated using stars[16], to allow us to assess whether there are differential effects. I will not discuss the "Simple" models (columns 1 and 2) in each case, since they are very similar (but for weighting) to the analysis from Tables 5 and 6 in Section 5.

## 7.1 School type

In the case of the proportion of applicants getting an offer, Table 7 shows no unexpected results when separating the successful proportion into those from independent and state schools. The only small deviation from this is that in 2008 the estimate for the proportion from independent schools is noticeably more negative than that for state schools (although still not statistically significant)[17]. However, this is not maintained in subsequent years and is reduced in the model with additional controls. This suggests that the introduction of the TSA has not had a differential overall impact on the proportion of all applicants who are ultimately offered a place and from come from each school type. However, this does not mean the same will be true at the intermediate stages of the process.

The additional controls in models 5 and 6 also behave as might be expected. There is a correlation between the mean number of GCSE A*s held by applicants of a given school type and the proportion of applicants who are successful and come from that same school type. We might also expect to see a negative relationship between average GCSE performance among one school type and the successful proportion from the other: to admissions tutors, applicants from different school types are substitutes and a rise in the performance of one of these groups might be expected to reduce demand for applicants from the other, other things being equal. However, if this effect exists it is too weak to be identified. The coefficients on the Norrington Score imply that a greater proportion of all applicants to colleges with higher performing existing undergraduates will be offered a place and come from state schools; there is no statistically significant effect on the proportion of all applicants who get an offer and come from an independent school. While the implications are rather difficult to interpret, its inclusion in the

[16] I conduct cross-model hypothesis testing using a seemingly-unrelated regression technique, specifically the Stata suest command, as this allows weights and clustering to be taken into account. Since the models being compared contain the same regressors this has no impact on the estimated standard errors (Zellner, 1962, p.351). Stars indicate statistical significance as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[17] Examining the results separately by PPE and E&M does not suggest this is driven by the relatively larger number of E&M interviews in that year.

Table 7: Proportion of all applicants getting an offer, comparing proportions who are successful and come from either or independent and state schools: difference in differences estimates

| Variable \ Model | Simple | | Years | | Controls | |
|---|---|---|---|---|---|---|
| | (1) Ind. | (2) State | (3) Ind. | (4) State | (5) Ind. | (6) State |
| Constant ($\alpha$) | 0.129 | 0.155 | 0.129 | 0.155 | 0.135 | -0.149 |
| | ( 0.006)*** | ( 0.006)*** | ( 0.006)*** | ( 0.006)*** | ( 0.156) | ( 0.147) |
| Treated ($\beta$) | -0.006 | -0.028 | -0.006 | -0.028 | -0.012 | -0.026 |
| | ( 0.012) | ( 0.012)** | ( 0.012) | ( 0.012)** | ( 0.011) | ( 0.009)*** |
| Policy On ($\gamma$) | -0.023 | -0.020 | | | | |
| | ( 0.004)*** | ( 0.004)*** | | | | |
| 2008 ($\gamma_8$) | | | -0.005 | -0.012 | -0.006 | -0.025 |
| | | | ( 0.004) | ( 0.004)*** | ( 0.010) | ( 0.009)*** |
| 2009 ($\gamma_9$) | | | -0.028 | -0.028 | -0.042 | -0.039 |
| | | | ( 0.004)*** | ( 0.005)*** | ( 0.006)*** | ( 0.006)*** |
| 2010 ($\gamma_{10}$) | | | -0.032 | -0.019 | -0.049 | -0.029 |
| | | | ( 0.006)*** | ( 0.005)*** | ( 0.008)*** | ( 0.006)*** |
| Treated*Policy On ($\delta$) | -0.008 | -0.005 | | | | |
| | ( 0.010) | ( 0.009) | | | | |
| Treated*2008 ($\delta_8$) | | | -0.026 | -0.004 | -0.020 | -0.005 |
| | | | ( 0.011)** | ( 0.012) | ( 0.012)* | ( 0.011) |
| Treated*2009 ($\delta_9$) | | | -0.005 | -0.013 | -0.005 | -0.012 |
| | | | ( 0.011) | ( 0.010) | ( 0.012) | ( 0.010) |
| Treated*2010 ($\delta_{10}$) | | | 0.005 | -0.001 | -0.007 | 0.003 |
| | | | ( 0.013) | ( 0.012) | ( 0.013) | ( 0.011) |
| Mean No. of GCSEs (State) | | | | | -0.021 | 0.010 |
| | | | | | ( 0.013) | ( 0.013) |
| Mean No. of GCSEs (Ind.) | | | | | -0.000 | -0.026 |
| | | | | | ( 0.010) | ( 0.010)** |
| Mean No. of A*s (State) | | | | | 0.003 | 0.015 |
| | | | | | ( 0.005) | ( 0.005)*** |
| Mean No. of A*s (Ind.) | | | | | 0.025 | -0.007 |
| | | | | | ( 0.005)*** | ( 0.004)* |
| Norrington Score / 10 | | | | | 0.477 | 6.254 |
| | | | | | ( 1.022) | ( 0.925)*** |
| **Differences in estimated effects by school type** | | | | | | |
| Treated*Policy On ($\delta$) | -0.003 | | | | | |
| Treated*2008 ($\delta_8$) | | | -0.022 | | -0.015 | |
| Treated*2009 ($\delta_9$) | | | 0.007 | | 0.007 | |
| Treated*2010 ($\delta_{10}$) | | | 0.006 | | -0.010 | |
| N | 116 | | 232 | | 232 | |

*Notes:* Analysis excludes individuals for whom school type is unknown. For Simple model (columns 1 and 2), Policy Off in 2005 and 2006; Policy On in 2008, 2009 and 2010. For other models (columns 3, 4, 5 and 6), base category for years is pooling of observations for 2005 and 2006. 'Ind.' is a contraction of Independent. Cross-model hypothesis testing conducted using seemingly-unrelated regressions. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Proportion of all applicants getting an interview, comparing proportions who are successful and come from either or independent and state schools: difference in differences estimates

| Variable \ Model | Simple | | Years | | Controls | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Ind. | State | Ind. | State | Ind. | State |
| Constant ($\alpha$) | 0.323 | 0.464 | 0.323 | 0.464 | 0.722 | -0.210 |
| | ( 0.015)*** | ( 0.015)*** | ( 0.015)*** | ( 0.015)*** | ( 0.331)** | ( 0.255) |
| Treated ($\beta$) | 0.065 | -0.024 | 0.065 | -0.024 | 0.053 | -0.018 |
| | ( 0.025)*** | ( 0.023)** | ( 0.025)*** | ( 0.023)** | ( 0.024)** | ( 0.021)*** |
| Policy On ($\gamma$) | -0.040 | -0.071 | | | | |
| | ( 0.009)*** | ( 0.009)*** | | | | |
| 2008 ($\gamma_8$) | | | -0.015 | -0.060 | -0.023 | -0.084 |
| | | | ( 0.008)** | ( 0.007)*** | ( 0.018) | ( 0.019)*** |
| 2009 ($\gamma_9$) | | | -0.041 | -0.078 | -0.063 | -0.090 |
| | | | ( 0.009)*** | ( 0.010)*** | ( 0.012)*** | ( 0.014)*** |
| 2010 ($\gamma_{10}$) | | | -0.060 | -0.073 | -0.092 | -0.084 |
| | | | ( 0.013)*** | ( 0.012)*** | ( 0.017)*** | ( 0.016)*** |
| Treated*Policy On ($\delta$) | -0.085 | -0.059 | | | | |
| | ( 0.016)*** | ( 0.018)*** | | | | |
| Treated*2008 ($\delta_8$) | | | -0.080 | -0.015 | -0.068 | -0.020 |
| | | | ( 0.017)*** | ( 0.017) | ( 0.020)*** | ( 0.020) |
| Treated*2009 ($\delta_9$) | | | -0.101 | -0.103 | -0.098 | -0.102 |
| | | | ( 0.021)*** | ( 0.023)*** | ( 0.024)*** | ( 0.023)*** |
| Treated*2010 ($\delta_{10}$) | | | -0.076 | -0.065 | -0.100 | -0.051 |
| | | | ( 0.021)*** | ( 0.025)*** | ( 0.020)*** | ( 0.027)* |
| Mean No. of GCSEs (State) | | | | | -0.019 | 0.030 |
| | | | | | ( 0.026) | ( 0.025) |
| Mean No. of GCSEs (Ind.) | | | | | -0.010 | -0.042 |
| | | | | | ( 0.025) | ( 0.018)** |
| Mean No. of A*s (State) | | | | | 0.001 | 0.021 |
| | | | | | ( 0.010) | ( 0.010)** |
| Mean No. of A*s (Ind.) | | | | | 0.054 | -0.024 |
| | | | | | ( 0.010)*** | ( 0.010)** |
| Norrington Score / 10 | | | | | -6.827 | 12.140 |
| | | | | | ( 2.072)*** | ( 1.983)*** |
| **Differences in estimated effects by school type** | | | | | | |
| Treated*Policy On ($\delta$) | -0.026 | | | | | |
| Treated*2008 ($\delta_8$) | | | -0.066** | | -0.048 | |
| Treated*2009 ($\delta_9$) | | | 0.003 | | 0.004 | |
| Treated*2010 ($\delta_{10}$) | | | -0.012 | | -0.048 | |
| N | 116 | | 232 | | 232 | |

*Notes:* Analysis excludes individuals for whom school type is unknown. For Simple model (columns 1 and 2), Policy Off in 2005 and 2006; Policy On in 2008, 2009 and 2010. For other models (columns 3, 4, 5 and 6), base category for years is pooling of observations for 2005 and 2006. 'Ind.' is a contraction of Independent. Cross-model hypothesis testing conducted using seemingly-unrelated regressions. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Proportion of all interviewees getting an offer, comparing proportions who are successful and come from either or independent and state schools: difference in differences estimates

| Variable \ Model | Simple | | Years | | Controls | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Ind. | State | Ind. | State | Ind. | State |
| Constant ($\alpha$) | 0.163 | 0.197 | 0.163 | 0.197 | 0.266 | -0.169 |
| | ( 0.007)*** | ( 0.007)*** | ( 0.007)*** | ( 0.007)*** | ( 0.196)** | ( 0.242) |
| Treated ($\beta$) | -0.016 | -0.043 | -0.016 | -0.043 | -0.027 | -0.041 |
| | ( 0.014)*** | ( 0.014)*** | ( 0.014)*** | ( 0.014)*** | ( 0.013)** | ( 0.012)*** |
| Policy On ($\gamma$) | -0.007 | 0.002 | | | | |
| | ( 0.005)*** | ( 0.005)*** | | | | |
| 2008 ($\gamma_8$) | | | 0.009 | 0.004 | -0.001 | -0.004 |
| | | | ( 0.005)* | ( 0.005)*** | ( 0.012) | ( 0.011)*** |
| 2009 ($\gamma_9$) | | | -0.014 | -0.007 | -0.037 | -0.021 |
| | | | ( 0.005)** | ( 0.006)*** | ( 0.008)*** | ( 0.008)** |
| 2010 ($\gamma_{10}$) | | | -0.015 | 0.011 | -0.046 | -0.003 |
| | | | ( 0.007)** | ( 0.007)*** | ( 0.010)*** | ( 0.009)*** |
| Treated*Policy On ($\delta$) | 0.016 | 0.018 | | | | |
| | ( 0.013)*** | ( 0.011)* | | | | |
| Treated*2008 ($\delta_8$) | | | -0.018 | 0.009 | -0.017 | 0.006 |
| | | | ( 0.014)*** | ( 0.016) | ( 0.014)*** | ( 0.015) |
| Treated*2009 ($\delta_9$) | | | 0.036 | 0.025 | 0.037 | 0.027 |
| | | | ( 0.018)* | ( 0.017)*** | ( 0.017)** | ( 0.018)*** |
| Treated*2010 ($\delta_{10}$) | | | 0.040 | 0.023 | 0.013 | 0.021 |
| | | | ( 0.019)** | ( 0.017)*** | ( 0.019)*** | ( 0.016)* |
| Mean No. of GCSEs (State) | | | | | -0.017 | -0.005 |
| | | | | | ( 0.016) | ( 0.017) |
| Mean No. of GCSEs (Ind.) | | | | | -0.010 | -0.012 |
| | | | | | ( 0.015) | ( 0.021)** |
| Mean No. of A*s (State) | | | | | 0.009 | 0.014 |
| | | | | | ( 0.007) | ( 0.009)** |
| Mean No. of A*s (Ind.) | | | | | 0.031 | -0.006 |
| | | | | | ( 0.006)*** | ( 0.007)** |
| Norrington Score / 10 | | | | | -1.354 | 7.320 |
| | | | | | ( 1.106)*** | ( 1.505)*** |
| **Differences in estimated effects by school type** | | | | | | |
| Treated*Policy On ($\delta$) | -0.002 | | | | | |
| Treated*2008 ($\delta_8$) | | | -0.027 | | -0.024 | |
| Treated*2009 ($\delta_9$) | | | 0.010 | | 0.011 | |
| Treated*2010 ($\delta_{10}$) | | | 0.017 | | -0.007 | |
| N | 116 | | 232 | | 231 | |

*Notes:* Analysis excludes individuals for whom school type is unknown. For Simple model (columns 1 and 2), Policy Off in 2005and 2006; Policy On in 2008, 2009 and 2010. For other models (columns 3, 4, 5 and 6), base category for years is pooling of observations for 2005 and 2006. 'Ind.' is a contraction of Independent. Cross-model hypothesis testing conducted using seemingly-unrelated regressions. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

model aims to help to control for the possibility that individuals attempt to choose colleges strategically to improve their chances of admissions.

Table 8 gives a more complex picture of the proportion of applicants who are called to interview: our simple DiD estimate was that the effect of the introduction of the aptitude test was more negative on the proportion of all applicants who were called to interview and came from independent schools than it was on the state school proportion, but that this difference was not statistically significant. However, from more flexible regression analysis we see that the estimated impact varies significantly year by year. Much of the difference in the simple estimates appears to be driven by a statistically significantly difference between the impacts by school type in 2008 ($\delta_8$)[18]. However, as with the proportion getting an offer, this difference between estimates becomes statistically insignificantly different from one another when we add controls to the model. Furthermore, by the following year this differential has vanished: in 2009 and 2010 the differences between the two estimates are in each case much smaller and not statistically significant. Considering the other controls in the model, there is also some evidence of a trade-off between candidates of different school types, with a positive effect of average GCSE performance of independent school applicants on the proportion of all applicants who get an offer and come from independent schools, but a negative effect of the same variable on the proportion from state schools. In summary, it would appear that any difference in effects may be driven by observable background characteristics, likely prior attainment, and is, at most, only short lived.

Finally, Table 9[19] also confirms the simple DiD estimates by failing to find strong evidence of a difference by school type in the proportion of interviewees who receive an offer. While there is (as with the proportion of applicants offered an interview) a noticeably larger difference by school type in 2008, it is not statistically significant. The inclusion of additional covariates makes a much smaller difference to the estimated effects (and the gap between them) than we saw in modelling the proportion of applicants offered an interview: this seems likely to be down to the smaller variation in observable characteristics between those interviewed.

[18]Examining these results separately for PPE and E&M (not reported here) suggests one of the reasons for this is that the policy setting a target number of interviews per place for E&M was not yet active. As such, the number of interviews for E&M weigh relatively larger than in other years. Focussing only on PPE, the estimate is for the same direction of difference in effects, but not statistically significant.

[19]The reduction in sample size in columns 5 and 6 in Table 9 is due to the fact that at one college in one year none of the state school applicants were invited to an interview.

The results from the regression analysis add confidence to findings from Section 5 in two ways. The estimates show a reasonably consistent story over time (particularly given the unusual circumstances in 2008); namely, that there is no evidence of different effects on the two proportions by school type. Second, they give some confidence that the results are not driven by changes in other observable characteristics, notably the average performance of applicants from each school type, or differences in college choice.

## 7.2 Gender

I now explore the results by gender in the same way. In the case of the proportion of applicants getting an offer, Table 10 confirms our earlier results. In no years are the differences by gender between the estimated effects statistically significant. As with analysis by school type, the additional controls in models 5 and 6 also behave as expected. There are positive correlations between the mean number of GCSE A*s held by applicants of a particular gender and the proportion of applicants who are successful and are of that gender. Likewise, any negative effects of increased performance by one gender on admissions chances of the other are either non-existent or too weak to be identified. The coefficients on the Norrington Score imply that a greater proportion of all applicants to colleges with higher performing existing undergraduates will be offered a place; this association is noticeably stronger for the success of male than female applicants, supporting its inclusion in the model.

Turning to the proportion of applicants called to interview, Table 11 shows a broadly consistent story of a larger decline in the proportion of applicants being called to interview who are male than the same proportion for females. However, the differences in estimated effects are not statistically significant. Examining these results separately for PPE and E&M (not reported here) suggests that the differences are driven more by changes in E&M. This seems likely to be because E&M received more applicants per place and, as such, the target number of interviews per place resulted in larger overall changes in the proportion of applicants called to interview[20]. Nevertheless, the results for PPE are not contradictory, but rather weaker.

Finally, Table 12 confirms the simple DiD estimate of a difference by gender in the proportion of all interviewees who receive an offer. The models find consistently statistically significant

[20]This is also hinted at by the smaller estimated effects in 2008, when this part of the policy had not yet been introduced for E&M.

Table 10: Proportion of all applicants getting an offer, comparing proportions who are successful and are either male or female: difference in differences estimates

| Variable \ Model | Simple | | Years | | Controls | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Female | Male | Female | Male | Female | Male |
| Constant ($\alpha$) | 0.135 | 0.149 | 0.135 | 0.149 | 0.087 | -0.076 |
| | ( 0.003)*** | ( 0.005)*** | ( 0.003)*** | ( 0.005)*** | ( 0.098) | ( 0.165) |
| Treated ($\beta$) | -0.046 | 0.012 | -0.046 | 0.012 | -0.053 | 0.008 |
| | ( 0.008)*** | ( 0.012) | ( 0.008)*** | ( 0.012) | ( 0.007)*** | ( 0.009) |
| Policy On ($\gamma$) | -0.020 | -0.022 | | | | |
| | ( 0.003)*** | ( 0.005)*** | | | | |
| 2008 ($\gamma_8$) | | | -0.007 | -0.010 | -0.010 | -0.016 |
| | | | ( 0.004) | ( 0.005)** | ( 0.006)* | ( 0.008)* |
| 2009 ($\gamma_9$) | | | -0.031 | -0.025 | -0.039 | -0.039 |
| | | | ( 0.004)*** | ( 0.006)*** | ( 0.004)*** | ( 0.006)*** |
| 2010 ($\gamma_{10}$) | | | -0.021 | -0.029 | -0.029 | -0.046 |
| | | | ( 0.005)*** | ( 0.005)*** | ( 0.005)*** | ( 0.007)*** |
| Treated*Policy On ($\delta$) | -0.013 | -0.000 | | | | |
| | ( 0.008)* | ( 0.012) | | | | |
| Treated*2008 ($\delta_8$) | | | -0.029 | -0.000 | -0.024 | -0.002 |
| | | | ( 0.011)*** | ( 0.015) | ( 0.009)*** | ( 0.014) |
| Treated*2009 ($\delta_9$) | | | -0.013 | -0.006 | -0.007 | -0.012 |
| | | | ( 0.008) | ( 0.012) | ( 0.008) | ( 0.011) |
| Treated*2010 ($\delta_{10}$) | | | -0.003 | 0.007 | -0.004 | 0.003 |
| | | | ( 0.009) | ( 0.015) | ( 0.009) | ( 0.014) |
| Mean No. of GCSEs (Male) | | | | | -0.008 | -0.033 |
| | | | | | ( 0.009) | ( 0.013)** |
| Mean No. of GCSEs (Female) | | | | | -0.007 | 0.006 |
| | | | | | ( 0.006) | ( 0.008) |
| Mean No. of A*s (Male) | | | | | -0.003 | 0.019 |
| | | | | | ( 0.004) | ( 0.005)*** |
| Mean No. of A*s (Female) | | | | | 0.020 | -0.002 |
| | | | | | ( 0.004)*** | ( 0.004) |
| Norrington Score / 10 | | | | | 1.334 | 5.987 |
| | | | | | ( 0.553)** | ( 1.237)*** |
| **Differences in estimated effects by gender** | | | | | | |
| Treated*Policy On ($\delta$) | -0.013 | | | | | |
| Treated*2008 ($\delta_8$) | | | -0.028 | | -0.021 | |
| Treated*2009 ($\delta_9$) | | | -0.007 | | 0.004 | |
| Treated*2010 ($\delta_{10}$) | | | -0.009 | | -0.007 | |
| N | 116 | | 232 | | 230 | |

*Notes:* Analysis excludes individuals for whom school type is unknown. For Simple model (columns 1 and 2), Policy Off in 2005 and 2006; Policy On in 2008, 2009 and 2010. For other models (columns 3, 4, 5 and 6), base category for years is pooling of observations for 2005 and 2006. Cross-model hypothesis testing conducted using seemingly-unrelated regressions. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Proportion of all applicants getting an interview, comparing proportions who are successful and are either male or female: difference in differences estimates

| Variable \ Model | Simple | | Years | | Controls | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Female | Male | Female | Male | Female | Male |
| Constant ($\alpha$) | 0.392 | 0.396 | 0.392 | 0.396 | 0.149 | 0.353 |
| | ( 0.009)*** | ( 0.011)*** | ( 0.009)*** | ( 0.011)*** | ( 0.214) | ( 0.261) |
| Treated ($\beta$) | -0.119 | 0.160 | -0.119 | 0.160 | -0.131 | 0.156 |
| | ( 0.018)*** | ( 0.022)*** | ( 0.018)*** | ( 0.022)*** | ( 0.012)*** | ( 0.019)*** |
| Policy On ($\gamma$) | -0.050 | -0.060 | | | | |
| | ( 0.008)*** | ( 0.009)*** | | | | |
| 2008 ($\gamma_8$) | | | -0.029 | -0.046 | -0.040 | -0.051 |
| | | | ( 0.008)*** | ( 0.009)*** | ( 0.012)*** | ( 0.015)*** |
| 2009 ($\gamma_9$) | | | -0.061 | -0.058 | -0.070 | -0.078 |
| | | | ( 0.009)*** | ( 0.011)*** | ( 0.008)*** | ( 0.011)*** |
| 2010 ($\gamma_{10}$) | | | -0.059 | -0.075 | -0.070 | -0.096 |
| | | | ( 0.010)*** | ( 0.010)*** | ( 0.009)*** | ( 0.013)*** |
| Treated*Policy On ($\delta$) | -0.057 | -0.087 | | | | |
| | ( 0.018)*** | ( 0.027)*** | | | | |
| Treated*2008 ($\delta_8$) | | | -0.048 | -0.047 | -0.036 | -0.051 |
| | | | ( 0.019)** | ( 0.028)* | ( 0.014)** | ( 0.024)** |
| Treated*2009 ($\delta_9$) | | | -0.082 | -0.122 | -0.070 | -0.131 |
| | | | ( 0.021)*** | ( 0.031)*** | ( 0.017)*** | ( 0.030)*** |
| Treated*2010 ($\delta_{10}$) | | | -0.049 | -0.092 | -0.046 | -0.100 |
| | | | ( 0.021)** | ( 0.028)*** | ( 0.015)*** | ( 0.027)*** |
| Mean No. of GCSEs (Male) | | | | | 0.028 | -0.048 |
| | | | | | ( 0.017)* | ( 0.025)* |
| Mean No. of GCSEs (Female) | | | | | -0.020 | -0.001 |
| | | | | | ( 0.012)* | ( 0.016) |
| Mean No. of A*s (Male) | | | | | -0.001 | 0.023 |
| | | | | | ( 0.008) | ( 0.008)*** |
| Mean No. of A*s (Female) | | | | | 0.025 | 0.003 |
| | | | | | ( 0.006)*** | ( 0.010) |
| Norrington Score / 10 | | | | | 0.183 | 5.830 |
| | | | | | ( 1.028)** | ( 1.899)*** |
| **Differences in estimated effects by gender** | | | | | | |
| Treated*Policy On ($\delta$) | 0.030 | | | | | |
| Treated*2008 ($\delta_8$) | | | -0.001 | | 0.015 | |
| Treated*2009 ($\delta_9$) | | | 0.041 | | 0.061 | |
| Treated*2010 ($\delta_{10}$) | | | 0.043 | | 0.054 | |
| N | 116 | | 232 | | 230 | |

*Notes:* Analysis excludes individuals for whom school type is unknown. For Simple model (columns 1 and 2), Policy Off in 2005 and 2006; Policy On in 2008, 2009 and 2010. For other models (columns 3, 4, 5 and 6), base category for years is pooling of observations for 2005 and 2006. Cross-model hypothesis testing conducted using seemingly-unrelated regressions. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Proportion of all interviewees getting an offer, comparing proportions who are successful and are either male or female: difference in differences estimates

| Variable \ Model | Simple | | Years | | Controls | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Female | Male | Female | Male | Female | Male |
| Constant ($\alpha$) | 0.172 | 0.188 | 0.172 | 0.188 | -0.047 | 0.010 |
| | ( 0.004)*** | ( 0.007)*** | ( 0.004)*** | ( 0.007)*** | ( 0.133) | ( 0.206) |
| Treated ($\beta$) | -0.064 | 0.006 | -0.064 | 0.006 | -0.068 | -0.002 |
| | ( 0.010)*** | ( 0.014)*** | ( 0.010)*** | ( 0.014)*** | ( 0.010)*** | ( 0.011)*** |
| Policy On ($\gamma$) | -0.002 | -0.002 | | | | |
| | ( 0.005)*** | ( 0.007)*** | | | | |
| 2008 ($\gamma_8$) | | | 0.008 | 0.006 | -0.001 | -0.003 |
| | | | ( 0.006)*** | ( 0.007)*** | ( 0.008)*** | ( 0.010)*** |
| 2009 ($\gamma_9$) | | | -0.017 | -0.004 | -0.028 | -0.026 |
| | | | ( 0.005)*** | ( 0.008)*** | ( 0.007)*** | ( 0.009)*** |
| 2010 ($\gamma_{10}$) | | | 0.002 | -0.007 | -0.011 | -0.030 |
| | | | ( 0.006)*** | ( 0.007)*** | ( 0.008)*** | ( 0.009)*** |
| Treated*Policy On ($\delta$) | -0.012 | 0.046 | | | | |
| | ( 0.011)*** | ( 0.016)*** | | | | |
| Treated*2008 ($\delta_8$) | | | -0.035 | 0.026 | -0.037 | 0.026 |
| | | | ( 0.015)** | ( 0.020)* | ( 0.013)*** | ( 0.018)** |
| Treated*2009 ($\delta_9$) | | | -0.006 | 0.067 | -0.016 | 0.064 |
| | | | ( 0.013)*** | ( 0.019)*** | ( 0.015)*** | ( 0.018)*** |
| Treated*2010 ($\delta_{10}$) | | | 0.003 | 0.060 | -0.011 | 0.051 |
| | | | ( 0.014)** | ( 0.023)*** | ( 0.015)*** | ( 0.023)** |
| Mean No. of GCSEs (Male) | | | | | 0.015 | -0.032 |
| | | | | | ( 0.012)* | ( 0.016)** |
| Mean No. of GCSEs (Female) | | | | | -0.013 | -0.004 |
| | | | | | ( 0.009)* | ( 0.019) |
| Mean No. of A*s (Male) | | | | | -0.001 | 0.022 |
| | | | | | ( 0.006) | ( 0.007)*** |
| Mean No. of A*s (Female) | | | | | 0.015 | -0.000 |
| | | | | | ( 0.005)*** | ( 0.007) |
| Norrington Score / 10 | | | | | 1.522 | 5.999 |
| | | | | | ( 0.915)* | ( 1.454)*** |
| **Differences in estimated effects by gender** | | | | | | |
| Treated*Policy On ($\delta$) | -0.058*** | | | | | |
| Treated*2008 ($\delta_8$) | | | -0.061** | | -0.063*** | |
| Treated*2009 ($\delta_9$) | | | -0.074*** | | -0.080*** | |
| Treated*2010 ($\delta_{10}$) | | | -0.057* | | -0.062** | |
| N | 116 | | 232 | | 230 | |

*Notes:* Analysis excludes individuals for whom school type is unknown. For Simple model (columns 1 and 2), Policy Off in 2005 and 2006; Policy On in 2008, 2009 and 2010. For other models (columns 3, 4, 5 and 6), base category for years is pooling of observations for 2005 and 2006. Cross-model hypothesis testing conducted using seemingly-unrelated regressions. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

evidence that the increase in the proportion of all interviewees receiving an offer is more positive for males than females. Generally this is explained by the increase in the proportion of all interviewees getting an offer being concentrated among males. Once again, the addition of covariates produces coefficients that conform to the pattern seen in earlier models. As with the results by school type, the inclusion of covariates in this model makes less difference than that seen for the earlier stage of the admissions process; however, if anything, their inclusion strengthens the statistical significance of the differences between the estimates for males and females.

I noted in Section 5 that an effect at the point of interview like this, given that the test is primarily used to screen applicants for interview, appears odd at first glance. However, a plausible explanation is that the TSA is more likely to screen out female applicants who would in the past have been offered a place once they were interviewed. Further investigation, considering combinations of gender and school type, suggests that this may be partly be due to a larger reduction in the proportion of all applicants invited to interview who were female and from an independent school. This is larger than the reduction in the proportion for the combination of female and state school. By contrast, the difference in effects between males and females from state schools in the proportion of all applicants getting an interview is much smaller. However, this only provides a potential pointer towards possible causes.

As with school type, the results from this regression analysis add confidence to findings from Section 5. When it comes to the proportion of interviewees who receive an offer, the regression estimates show a consistent and statistically significant set of estimates over time, with the overall increases driven by the proportion who receive an offer and are male. Furthermore, the regression models with additional controls suggest that the results are not driven by changes in other observable characteristics within the groups.

# 8 Alternative outcome measures

Proportions of applicants who are successful and come from a particular gender or school type is not the only way to think about the admissions process. In this section, I take an alternative approach, looking at each stage of the admissions process and analysing the share of the individuals that come from each of our groups of interest. Since all applicants in the dataset are classified as coming from either independent or state schools, the shares of each sum to 1. The same is the case for males and females. As such, we can restrict interest to just one of the shares in each case: I choose the share who come from a state school and the share who are female. Returning to the graphical representation of the admissions process in Figure 1, instead of considering the decision points themselves, I analyse the share of applicants, interviewees, and those who receive an offer who come from state schools and, separately, the share who are female.

Concentrating on outcome measures of this type, generally with respect to school type, is popular in the press (for example Vasagar, 2011), perhaps because a single figure is more readily comprehensible. Furthermore, while the main analysis produced estimated effects that are comparable in absolute terms, this alternative approach implicitly takes into account the size of the effects relative to the baseline proportion of successful applicants of each type. We will see the importance of this in the discussion of the results by gender below.

This alternative approach also allows us to consider an important additional aspect, which the main analysis was not able to address. As discussed in Section 3, the proportion of young people who choose to apply cannot be analysed, since potential applicants are not observed by the University. However, a related, though not identical, question is whether there is an impact on the make up of the pool of applicants i.e. the share of applicants who are female, or the share from state schools. An increase in the proportion of applicants from independent schools who do in fact apply will decrease this figure (holding state school applications constant) and vice versa. Rather than taking as a given the pool of applicants or interviewees, as the main analysis does, this approach focuses on the cumulative effect of the policy change (including changes in application behaviour) up to a given point in the admissions process. One drawback of these outcome variables is that they do not tell us about any overall changes in the number of interviews and offers.

Turning to school type first, I apply the same DiD method as for the analysis in Section 5 to identify the impact of the introduction of the TSA on the relative numbers of applicants from independent and state schools by comparing the change in share of applicants, interviewees and those receiving an offer between Economics and other subjects[21]. Adopting the same

[21] I do subject these figures to the same regression analysis as used above, but do not report these results.

notation as that introduced in Section 5 the outcome variables are as follows:

$$\text{Share of } \textbf{applicants} \text{ from state schools: } \frac{A_S}{A_I + A_S}$$
$$\text{Share of } \textbf{interviewees} \text{ from state schools: } \frac{I_S}{I_I + I_S}$$
$$\text{Share of those } \textbf{offered} \text{ a place from state schools: } \frac{O_S}{O_I + O_S}$$

How do these relate to the outcome variables for my main analysis? While those took the form $\frac{I_S}{A_I+A_S}$ (in the case of the proportion of all applicants called to interview and coming from a state school), these alternative outcome variables concentrate on proportions within a particular stage of the admissions process. They have the same denominators as the main analysis's outcomes, but quite different numerators.

Table 13: Share of applicants from State schools, share of interviewees from State schools, and share of those who receive an offer from State schools, by year and subject group: simple difference in differences estimates

| **Applicants** | **Policy Off** | **Policy On** | **Difference** |
|---|---|---|---|
| Economics | 0.551 | 0.575 | 0.024 |
| | ( 0.023) | ( 0.023) | ( 0.014)* |
| Others | 0.617 | 0.632 | 0.015 |
| | ( 0.016) | ( 0.013) | ( 0.008)* |
| Difference | -0.066 | -0.057 | 0.009 |
| | ( 0.028)*** | ( 0.026)** | ( 0.016) |
| **Interviewees** | **Policy Off** | **Policy On** | **Difference** |
| Economics | 0.527 | 0.536 | 0.009 |
| | ( 0.023) | ( 0.026) | ( 0.013) |
| Others | 0.592 | 0.582 | -0.010 |
| | ( 0.018) | ( 0.014) | ( 0.009) |
| Difference | -0.066 | -0.046 | 0.020 |
| | ( 0.029)** | ( 0.029) | ( 0.016) |
| **Offered** | **Policy Off** | **Policy On** | **Difference** |
| Economics | 0.508 | 0.527 | 0.019 |
| | ( 0.032) | ( 0.025) | ( 0.026) |
| Others | 0.548 | 0.561 | 0.013 |
| | ( 0.017) | ( 0.014) | ( 0.011) |
| Difference | -0.040 | -0.034 | 0.006 |
| | ( 0.036) | ( 0.029) | ( 0.028) |

*Notes:* Analysis excludes individuals for whom school type is unknown. Policy Off in 2005, 2006 and 2007; Policy On in 2008, 2009 and 2010. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate statistical significance: $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$. Sample sizes: Applicants: 63986 Interviewees: 46106 Attendees: 16412

By reading across the rows in the top panel of Table 13, we can see that the share of applicants from state schools has been rising in all subjects, Economics included. Figure 4 shows a large increase in the number of applications from state schools, suggesting this is the cause, rather than any decline in the number of applications from independent schools. Furthermore, the difference between Economics and other subjects (seen by reading down each column) shows that Economics applicants are more likely to be from independent schools than those to other subjects. However, the DiD estimate, in the bottom right hand cell, highlights that the increase was not statistically significantly larger in Economics when the TSA was introduced: there is no strong evidence that the introduction of the TSA affected the makeup of applicants in this way. It should be remembered that this analysis only covers the three years following the introduction of the policy; changes in behaviour by applicants are likely to take some time.

Unlike among applicants, there is only a very small rise in the proportion of Economics interviewees who come from state schools. In fact, among non-Economics subjects the proportion declines a small amount, however this is far from statistical significance. With no significant changes in the proportion of interviewees from state school among either the treatment or control groups it comes as little surprise that the DiD estimate provides no evidence of a statistically significant effect of the policy on the proportion of interviewees who come from a state school.

Finally, considering the proportion of those offered a place that come from state schools (the statistic that receives most popular attention), the story is very similar to that for interviewees. In each case, these results echo the findings from Section 5, suggesting that the policy does not have a large impact on the kinds of young people who make it through the admissions process.

Subjecting the analysis in this section to the same regression modelling as in Section 6 does not materially alter the interpretation of these findings. I also take the approach further in analysing differences by socioeconomic status in Appendix A, using the applicants' schools' IDACI (Income Deprivation Affecting Children and Infants Index) figure as the outcome of interest. The analysis does not seem inconsistent with the findings reported above.

Turning now to the same analysis by gender, the story seems initially similar. The DiD estimate of the effect on the share of applicants who are female is zero. However, there is change in the composition of interviewees. The share of interviewees for Economics who are female falls by 3.6 percentage points, at a time when this figure is rising (marginally) among other subjects. This results in an estimated impact of the TSA of a 4.5 percentage point reduction

Table 14: Share of applicants who are female, share of interviewees who are female, and share of those who receive an offer who are female, by year and subject group: simple difference in differences estimates

| **Applicants** | **Policy Off** | **Policy On** | **Difference** |
|---|---|---|---|
| Economics | 0.325 | 0.323 | -0.002 |
| | ( 0.013) | ( 0.008) | ( 0.014) |
| Others | 0.505 | 0.502 | -0.003 |
| | ( 0.013) | ( 0.009) | ( 0.012) |
| Difference | -0.180 | -0.179 | 0.000 |
| | ( 0.018)*** | ( 0.011)*** | ( 0.018) |
| **Interviewees** | **Policy Off** | **Policy On** | **Difference** |
| Economics | 0.326 | 0.289 | -0.036 |
| | ( 0.014) | ( 0.011) | ( 0.016)* |
| Others | 0.497 | 0.505 | 0.009 |
| | ( 0.011) | ( 0.009) | ( 0.010) |
| Difference | -0.171 | -0.216 | -0.045 |
| | ( 0.018)*** | ( 0.014)*** | ( 0.019)*** |
| **Offered** | **Policy Off** | **Policy On** | **Difference** |
| Economics | 0.355 | 0.293 | -0.061 |
| | ( 0.019) | ( 0.018) | ( 0.027)* |
| Others | 0.476 | 0.478 | 0.002 |
| | ( 0.012) | ( 0.011) | ( 0.013) |
| Difference | -0.122 | -0.184 | -0.063 |
| | ( 0.022)*** | ( 0.021)*** | ( 0.029)** |

*Notes:* Analysis excludes individuals for whom school type is unknown. Policy Off in 2005, 2006 and 2007; Policy On in 2008, 2009 and 2010. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate statistical significance: $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$. Sample sizes: Applicants: 63986 Interviewees: 46106 Attendees: 16412 in the share of interviewees who are female. Furthermore, regression analysis (allowing for different effects by year and including the same covariates as in the main analysis) casts little doubt on this finding.

Why do these results seemingly differ from our findings for gender in the main analysis, where we saw the proportion of applicants offered an interview and who are male decline more than the proportion of all applicants offered an interview and who are female? It is because the proportion for males starts at a higher baseline than for females; as such, the larger absolute decline for the male proportion has a relatively smaller effect on the gender makeup of interviewees.

Considering those offered a place the figures are similar: there is a fall in the female share of those offered a place to study Economics, despite the opposite trend among other subjects. This leads to an estimated negative effect of the TSA of 6.3 percentage points. However, unlike in the case of interviewees, these estimates are reduced to statistical insignificance by the inclusion of additional controls in regression analysis.

These results do not suggest that the introduction of the TSA has had a detrimental effect on the proportion of female applicants to Economics courses at the University of Oxford. However, a gap would appear to open in the share of interviewees who are female, and hence on into the share of those offered a place. The estimated effects are larger than those we recovered above for changes in shares from state schools. However, in this case, regression analysis reduces rather than adds to our confidence: the statistical evidence is only remains strong in the case of the share of interviewees who are female.

# 9 Robustness

The extent to which we can trust the findings from DiD analysis rests on the validity of the common trends assumption that underlies it. This cannot be tested directly, since the trend we would wish to look at is an unobserved counterfactual. However, robustness checks can provide some evidence that the assumption seems likely to hold.

The first of these I employ is a ‘placebo’ test. This involves estimating the effect across a period when the policy was not introduced, in this case between 2005 and 2006. The treatment and control groups remain as specified for the main analysis (Economics as treatment, all other subjects as controls). Finding an effect during this period, when there was no policy to produce one, would suggest a failure of the common trends assumption was inducing the apparent impact. The results from the placebo treatment on the proportion of all applicants who get a place, all applicants who get an interview and all interviewees who get a place are shown in Table 15, using the same output from linear regression employed in Section 7. No significant effect is identified at any stage of the admissions process, which is reassuring. This continues to hold true when the proportions of applicants are analysed separately by school type or gender (not shown).

Table 15: Proportion of all applicants getting an offer, all applicants getting an interview, and all interviewees getting an offer - placebo test: difference in differences estimates

| | (1) | (2) | (3) |
|---|---|---|---|
| | Offer | Interview | Inter.→Offer |
| Constant ($\alpha$) | 0.292 | 0.805 | 0.362 |
| | (0.006)$^{***}$ | (0.007)$^{***}$ | (0.007)$^{***}$ |
| Treated ($\beta$) | -0.040 | 0.050 | -0.066 |
| | (0.016)$^{**}$ | (0.016)$^{***}$ | (0.019)$^{***}$ |
| Policy Placebo ($\gamma$) | -0.014 | -0.033 | -0.003 |
| | (0.006)$^{**}$ | (0.004)$^{***}$ | (0.007) |
| Treated*Policy Placebo ($\delta$) | 0.013 | -0.012 | 0.017 |
| | (0.016) | (0.021) | (0.018) |
| $N$ | 116 | 116 | 116 |
| $R^2$ | 0.064 | 0.157 | 0.128 |

*Notes:* Analysis excludes individuals for whom school type is unknown. Policy Off in 2005; Policy On in 2006. Standard errors, clustered by college, in parentheses. Stars indicate staistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Second, I alter my control group to one which should even more closely resemble the treatment group: applicants to Social Science courses[22]. Table 16 shows the results, with the interaction between Economics and Policy On ($\delta$) being the key coefficient of interest in each model. It shows the estimated impact on the proportion of applicants getting an interview as being a reduction of 22.9 percentage points, while for the proportion of interviewees getting a place the estimate is an increase of 6.0 percentage points. These are qualitatively similar to the estimates in the main analysis of 14.4 percentage points and 6.4 percentage points, respectively. The impact on the proportion of applicants who get a place is estimated at close to zero

[22]I define Social Science courses as follows: Experimental Psychology; Geography; History and Economics (although an Economics subject this did not introduce the TSA); History and Politics; Law; Law with Law Studies in Europe; and Psychology, Philosophy and Physiology (PPP).

Table 16: Proportion of applicants getting an offer, applicants getting an interview, and interviewees getting an offer - restricted control group: difference in differences estimates

| | (1) Offer | (2) Interview | (3) Inter.$\rightarrow$Offer |
|---|---|---|---|
| Constant ($\alpha$) | 0.245 | 0.667 | 0.368 |
| | (0.007)$^{***}$ | (0.014)$^{***}$ | (0.010)$^{***}$ |
| Treated ($\beta$) | 0.005 | 0.162 | -0.066 |
| | (0.016) | (0.019)$^{***}$ | (0.020)$^{***}$ |
| Policy On ($\gamma$) | -0.031 | -0.050 | -0.016 |
| | (0.007)$^{***}$ | (0.012)$^{***}$ | (0.011) |
| Treated*Policy On ($\delta$) | -0.025 | -0.204 | 0.046 |
| | (0.014)$^{*}$ | (0.025)$^{***}$ | (0.021)$^{**}$ |
| $N$ | 116 | 116 | 116 |
| $R^2$ | 0.148 | 0.597 | 0.108 |

*Notes:* Analysis excludes individuals for whom school type is unknown. Policy Off in 2005, 2006 and 2007; Policy On in 2008, 2009 and 2010. Standard errors, clustered by college, in parentheses. Stars indicate staistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 17: Proportion of all applicants getting an offer, an interview, and interviewees getting an offer - comparing applicants from schools in high and low SES areas: difference in differences estimates

| Variable \ Outcome | Offer | | Interview | | Interview$\rightarrow$Offer | |
|---|---|---|---|---|---|---|
| | (1) High | (2) Low | (3) High | (4) Low | (5) High | (6) Low |
| Constant ($\alpha$) | 0.140 | 0.150 | 0.363 | 0.429 | 0.177 | 0.190 |
| | (0.004)$^{***}$ | (0.004)$^{***}$ | (0.007)$^{***}$ | (0.007)$^{***}$ | (0.005)$^{***}$ | (0.005)$^{***}$ |
| Treated ($\beta$) | -0.015 | -0.020 | 0.055 | -0.013 | -0.027 | -0.034 |
| | (0.011) | (0.011)$^{*}$ | (0.018)$^{***}$ | (0.017) | (0.013)$^{**}$ | (0.013)$^{**}$ |
| Policy On ($\gamma$) | -0.020 | -0.023 | -0.040 | -0.068 | -0.002 | -0.004 |
| | (0.004)$^{***}$ | (0.003)$^{***}$ | (0.005)$^{***}$ | (0.004)$^{***}$ | (0.004) | (0.004) |
| Treated*Policy On ($\delta$) | -0.004 | -0.010 | -0.094 | -0.050 | 0.023 | 0.013 |
| | (0.011) | (0.010) | (0.018)$^{***}$ | (0.019)$^{**}$ | (0.014) | (0.014) |
| $N$ | 116 | 116 | 116 | 116 | 116 | 116 |
| $R^2$ | 0.137 | 0.218 | 0.440 | 0.456 | 0.058 | 0.092 |

*Notes:* Analysis excludes individuals for whom school type is unknown. Policy Off in 2005, 2006 and 2007; Policy On in 2008, 2009 and 2010. Standard errors, clustered by college, in parentheses. Stars indicate staistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

and statistically insignificant. Once again, there is little divergence from this picture when the proportions of applicants are analysed separately by school type or gender (not shown).

Finally, I employ an alternative proxy of socioeconomic status. Instead of attendance at an independent school, I define a binary variable set to zero when applicants attend schools in the three most deprived fifths of postcodes, according to the Index of Deprivation Affecting Children and Infants (IDACI)[23], and set to one when they attend schools in the least two deprived fifths of postcodes. This roughly replicates the proportions of independent school applicants. The polychoric correlation between an individual attending an independent school and attending a school in a 'high SES area' is 0.37. Looked at another way, 52% of individuals in the dataset who attend a school in a 'high SES area' are attending an independent school. By contrast, only 29% of those attending a school in a 'low SES area' are attending an independent school. I re-estimate my DiD model, with successful proportions split by this variable.

The results are shown in Table 17 and produce qualitatively similar estimates to those from the main analysis. For example, the proportion of all applicants who are called to interview and come from a school in a high SES area is reduced by 7.9 percentage points, compared with 8.5 percentage points for independent schools. Similarly, the proportion of all applicants who are called to interview and come from a school in a low SES area is reduced by 5.0 percentage points, compared with 5.9 percentage points for state schools.

The results from these robustness checks are very encouraging, producing no significant effect from a placebo test and substantively similar results to my main analysis for the two other tests.

# 10 Conclusions

This paper has estimated the effects of introducing an aptitude test to an elite university's admissions process using difference in differences methods and data from the University of Oxford. No evidence is found of an overall impact on the proportion of applicants who receive an offer of a place to study at the University. The policy was coupled with a policy setting

[23] I take an alternative approach to analysis using IDACI in Appendix A. This does not involve converting it to a dichotomous variable in this way, which does reduce the informative content of the variable. I also include more detail on the construction of the IDACI.

a target number of interviews per place, reducing the proportion of applicants invited to interview (by 14 percentage points). Offsetting this, the proportion of interviewees receiving an interview increased (by 3.6 percentage points), driven by the reduction in the number of interviewees rather than an increase in the number of offers.

There is no clear evidence of differential effects on the proportion of all applicants offered a place by the school type individuals come from. Splitting the admissions process into its constituent parts: at first glance, there appeared to be evidence that the reduction in the proportion of applicants called to interview had a larger (negative) effect on the proportion of all applicants getting an interview who come independent school, although when examined more closely this was driven by peculiarities relating to the first year of introduction. Furthermore, there is little convincing evidence of heterogeneity by school type in the proportion of interviewees offered a place.

In the case of differences by gender, while there no strong evidence of overall differences between the effects on the proportion of all applicants getting an offer and who come from each gender, there is some evidence of males and females being affected differently by the introduction of an aptitude test at different points of the admissions process. Males appear relatively less likely to be called for an interview, while female interviewees are subsequently less likely to be offered a place. However, the statistical evidence is weaker in the case of the former.

To return to the question posed in the title, I do not find strong evidence that introducing an aptitude test to the admissions process of an elite university will have differing effects on applicants' chances of being offered a place depending on their socioeconomic status. Furthermore, while I do find differences in the effects of introducing the test on each gender at different points of the admissions process, I do not find strong evidence that the introduction of an aptitude test affects the relative chances of admission by sex.

# Bibliography


Admissions Testing Service (2013a). About TSA Cambridge. Available from: http://www.admissionstestingservice.org/our-services/thinking-skills/tsa-cambridge/about-tsa-cambridge/ and retrieved on 25/09/2013.

Admissions Testing Service (2013b). Assessments in thinking skills. Available from: http://www.admissionstestingservice.org/our-services/thinking-skills/ and retrieved on 25/09/2013.

Anders, J. (2012a). The Link between Household Income, University Applications and University Attendance. *Fiscal Studies*, 33(2):185–210.

Anders, J. (2012b). Using the Longitudinal Study of Young People in England for research into Higher Education access. DoQSS Working Paper 12-13, Institute of Education, University of London.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics*. Princeton University Press, Princeton.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust difference-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275.

Bhattacharya, D., Kanaya, S., and Stevens, M. (2012). Are University Admissions Academically Fair? Department for Economics Working Paper, University of Oxford, Oxford.

Boliver, V. (2013). How fair is access to more prestigious UK universities? *British Journal of Sociology*, 64(2):344–364.

Brewer, M., Crossley, T. F., and Joyce, R. (2013). Inference with difference-in-differences revisited. IZA Discussion Paper 7742, Institute for the Study of Labor.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427.

Card, D. (1992). Using regional variation in wages to measure the effects of the federal minimum wage. *Industrial and Labor Relations Review*, 46(1):22–37.

Chevalier, A. and Conlon, G. (2003). Does it pay to attend a prestigious university? CEE Discussion Paper 33, Centre for the Economics of Education, London School of Economics.

Department for Education (2010). Schools, Pupils, and their Characteristics, January 2010. Statistical First Release SFR 09/2010, Department for Education, London.

Emery, J. L. (2006). Predicting degree performance with the Thinking Skills Assessment: Report 2. Research report, Cambridge Assessment, Cambridge.

Emery, J. L., Bell, J. F., and Shannon, M. D. (2006). Predicting degree performance with the Thinking Skills Assessment. Research report, Cambridge Assessment, Cambridge.

Geiser, S. (2008). Back to the Basics: In Defense of Achievement (and Achievement Tests) in College Admissions. CSHE Research and Occational Paper CSHE.12.08, Centre for Studies in Higher Education, UC Berkeley, Berkley.

Kennedy, P. (2008). *A Guide to Econometrics*. Blackwell Publishing, Oxford, 6th edition.

Kirkup, C., Wheater, R., Morrison, J., Durbin, B., and Pomati, M. (2010). Use of an Aptitude Test in University Entrance: A Validity Study. BIS Research paper 26, Department for Business, Innovation and Skills, London.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Linn, M. C. and Hyde, J. S. (1989). Gender, mathematics, and science. *Educational Researcher*, 18:17–27.

McDonald, A. S., Newton, P. E., and Whetton, C. (2001a). A Pilot of Aptitude Testing for University Entrance. Research report, National Foundation for Educational Research, Slough.

McDonald, A. S., Newton, P. E., Whetton, C., and Benefield, P. (2001b). Aptitude Testing for University Entrace: A Literature Review. Research report, National Foundation for Educational Research, Slough.

McLennan, D., Barnes, H., Noble, M., Davies, J., and Garratt, E. (2011). The English Indices of Deprivation 2010. Department for Communities and Local Government technical report.

Messick, S. (1989). Meaning and values in test validation: the science and ethics of assessment. *Education Researcher*, 2(18):5–11.

Office of National Statistics (2014). Super Output Areas. Available from: http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas--soas-/index.html and retrieved on 20/01/2014.

Pallais, A. and Turner, S. (2006). Opportunities for Low-Income Students at Top Colleges and Universities: Policy Initiatives and the Distrbution of Students. *National Tax Journal*, 59(2):357–386.

Pallais, A. and Turner, S. (2008). Access to Elites. In Dickert-Conlin, S. and Rubenstein, R., editors, *Economic Inequality and Higher Education: Access, Perstistence, and Success*. Russell Sage Foundation, London.

Parente, P. M. D. C. and Santos Silva, J. M. C. (2013). Quantile regression with clustered data. Department of Economics Discussion Paper 728, University of Essex.

Rothstein, J. (2002). College Performance Predictions and the SAT. Working Paper 45, Center for Labor Economics, University of California Berkeley.

Stringer, N. (2008). Aptitude tests versus school exams as selection tolls for higher education and the case for assessing educational achievement in context. *Research Papers in Education*, 23(1):53–68.

Tannenbaum, D. I. (2012). Do gender differences in risk aversion explain the gender gap in SAT scores? Uncovering risk attitudes and the test score gap. Working paper, Department of Economics, University of Chicago.

University of Oxford (2013). Undergraduate Degree Classifications 2012/13. Available from: http://www.ox.ac.uk/about_the_university/facts_and_figures/norringtontable.html and retrieved on 25/09/2013.

University of Oxford (2014). Use of contextual data. Available from: http://j.mp/Ppvdv7 (University of Oxford website) and retrived on 24/02/2014.

Vasagar, J. (2011). Oxford University set for record state school intake. The Guardian. Available from: http://gu.com/p/2ny4z/tw and retrieved on 25/09/2013.

Wikström, M. and Wikström, C. (2014). Who benefits from university admissions tests? - a comparison between grades and test score as selection instruments to higher education. Umeå Economic Studies 874, Department of Economics, Umeå University.

Willetts, D. (2011). Fair Access to Higher Education - Letter to the Director of Fair Access, and the National Scholarship Programme. Written Ministerial Statement, Department for Business Innovation and Skills, London.

Wilmouth, D. (1991). Should the SAT be a Factor in College Admissions. ED 345 592, U.S. Department of Education, Educational Resources Information Centre, Washington, DC.

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association*, 57(298):348–368.

Zwick, R. (2007). College Admission Testing. Report, National Association for College Admission Counselling.

# A Effects on an area-level deprivation index

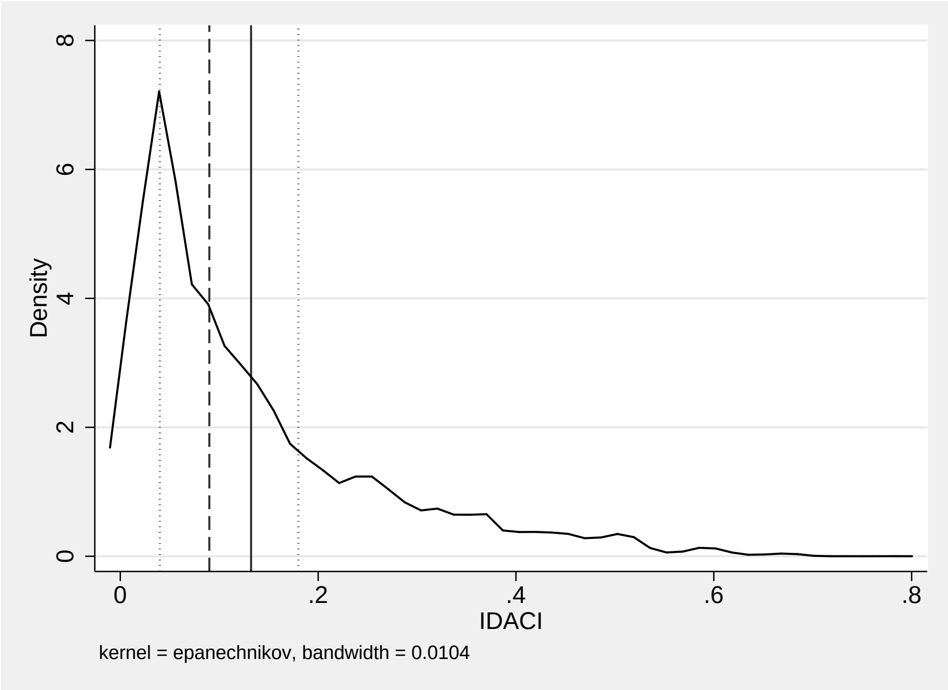
Using the same approach to analysing stages of the admissions process as that used in Section 8, I also consider the effect of introducing the TSA on another proxy for applicants' SES. I use the average area deprivation level of applicants' schools, measured using the Income Deprivation Affecting Children Index (IDACI) that I described in Section 4.

The IDACI is constructed as the percentage of all children aged 0-15 living in income deprived families (McLennan et al., 2011, p.22-23) within a Lower Layer Super Output Area (geographical districts covering the UK containing between 400 and 1,200 households (Office of National Statistics, 2014)). This is reported to the nearest whole percent. Nevertheless, it gives more potential discrimination than the simple independent/state split used in my main analysis. Figure 6 shows the graph of a kernel density estimate of the school IDACI of individuals in the dataset. It shows that the distribution is highly skewed, with applicants to the University of Oxford highly concentrated in schools in low-deprivation areas. This is also reflected in the difference between the mean (13%) and the median (9%). Unfortunately, school IDACI is missing in more cases (11.1%) than school type (2.2%): 11.4% of applicants at independent schools, 6.5% of applicants at state schools, and 83.4% of applicants with some other or missing school type have no school IDACI recorded.

While it would be better to use the IDACI for the young person's area of residence (rather than that of their school), this was not available for reasons of confidentiality. However, analysis using the Longitudinal Study of Young People in England (following a cohort of roughly similar age to those in the administrative data) shows that the IDACI score of a young person's school's area is correlated with their own socioeconomic status. I report the results in Table 18. The correlation between the IDACI score for the area where a young person lives is positively correlated with the IDACI score of the area where their school is situated (Pearson's correlation coefficient = 0.46). More fundamentally, the IDACI score of the area where a young person's school is situated is weakly negatively correlated (since one is a measure of disadvantage and the other a measure of advantage) with their household income (Pearson's correlation coefficient = -0.21).

Using a continuous outcome variable also allows analysis of changes to different parts of the distribution of applicants' schools' area deprivation, not just changes to the mean. Although

Figure 6: Kernel density distribution of IDACI score




**Notes:** Solid vertical line indicates mean, dashed vertical line indicates median, and dotted vertical lines indicate upper and lower quartiles. Excludes individuals for whom school IDACI was not recorded.

Table 18: Average characteristics of Longitudinal Study of Young People in England cohort members by IDACI quintile group of their school's area

| | **IDACI quintile group of school's area** | | | | |
|---|---|---|---|---|---|
| **Characteristic** | 5th (Advantaged) | 4th | 3rd | 2nd | 1st (Disadvantaged) |
| IDACI score of young person's home area (%) | 15 | 18 | 23 | 28 | 39 |
| Household Income (£) | 22,579 | 21,355 | 18,017 | 17,158 | 14,233 |
| Mother has a degree (%) | 30 | 26 | 22 | 20 | 14 |
| Father has higher managerial or professional occupation (%) | 43 | 39 | 31 | 29 | 20 |
| Family in financial difficulties (%) | 6 | 6 | 7 | 9 | 11 |
| Family living in socially rented housing (%) | 15 | 18 | 22 | 29 | 41 |
| Young person attends independent school (%) | 6 | 5 | 0 | 5 | 0 |

*Notes:* Data from the Longitudinal Study of Young People in England (LSYPE). Average characteristics for LSYPE cohort members who attend schools in each of five quintiles groups defined by the IDACI score of the school's area. Characteristics are measured at Wave 1 of the LSYPE, at age 14 years, except in case of income, which is averaged over measurements are ages 14, 15 and 16. Income is in 2003–2004 prices. Calculations courtesy of Claire Crawford of the Institute for Fiscal Studies/University of Warwick.

the method I use is not quantile regression (Koenker and Bassett, 1978; Parente and Santos Silva, 2013), it shares some of the same intuition. As in earlier sections of the paper I use college-level least squares regression, but rather than only using as observations the mean deprivation level of applicants (or interviewees, or those offered a place), I also use models with observations constructed as the lower quartiles (Q25), medians or upper quartiles (Q75) of the school IDACI for a given college, course, year combination.

Such changes are matters of interest since a shift in the mean deprivation level alone could result from a number of different changes in the underlying distribution of applicants, interviewees or those offered a place. To illustrate this, let us consider two notional shifts in the deprivation distribution of interviewees which could have identical effects on the mean deprivation of applicants. We might see an effect that only shifts the lower quartile of the deprivation distribution of interviewees and has no impact on the median or the upper quartile. This would suggest that the policy change is filtering out some of the applicants from most advantaged schools, but these are being replaced by applicants only slightly above them on the deprivation distribution. The effect is not having a broader impact further up the distribution. Alternatively, we might see an effect that shifts the lower quartile of the distribution of interviewees somewhat less than our first change, but also shifts the median interviewee's deprivation level. This would imply a somewhat broader effect, with those at the bottom of the deprivation distribution being replaced by applicants significantly further down (albeit without much effect on those attending schools in the most deprived areas).

I report the results from regression models similar to those from Section 6, with the coefficient on the interaction between the policy on and treatment group ($\delta$) recovering the DiD estimate, for each stage of the admissions process in Tables 19, 20 and 21. The estimates of the policy are in units of the IDACI. For example, an estimate of 1 implies an estimated 1 percentage point increase in the mean, median or quartile deprivation of applicants, interviewees or those offered a place. As such, their magnitudes are not comparable with estimates in Section 8. As with the main analysis, I include controls for the average GCSE performance by state and independent school applicants, interviewees or attendees and college Norrington score.

We see from Table 19, in common with the analysis in Section 8, no statistically significant estimated effect on the mean IDACI of applicants' schools. If anything, the results estimate an increase in the mean area deprivation level of applicants' schools equivalent to 3 additional

Table 19: School IDACI of applicants - changes at the mean, lower quartile, median and upper quartile of colleges' distributions: difference in differences estimates

| | (1) Mean | (2) Q25 | (3) Median | (4) Q75 |
|---|---|---|---|---|
| Constant ($\alpha$) | 5.997 | 7.754 | 6.175 | -25.776 |
| | (9.765) | (5.248) | (10.976) | (28.401) |
| Treated ($\beta$) | -0.332 | -0.085 | -0.220 | 0.324 |
| | (0.355) | (0.229) | (0.407) | (0.729) |
| Policy On ($\gamma$) | 0.679 | 0.397 | 0.567 | 0.581 |
| | (0.381)$^{*}$ | (0.191)$^{**}$ | (0.482) | (1.064) |
| Treated*Policy On ($\delta$) | 0.333 | 0.131 | 0.260 | 0.048 |
| | (0.422) | (0.227) | (0.445) | (0.933) |
| Mean No. of GCSEs (State) | 0.583 | -0.638 | -0.733 | 5.009 |
| | (0.966) | (0.364)$^{*}$ | (1.228) | (2.722)$^{*}$ |
| Mean No. of GCSEs (Ind.) | -0.076 | 0.481 | 0.930 | -0.625 |
| | (0.616) | (0.366) | (0.588) | (1.676) |
| Mean No. of A*s (State) | -0.348 | 0.344 | 0.511 | -2.649 |
| | (0.222) | (0.138)$^{**}$ | (0.326) | (0.952)$^{***}$ |
| Mean No. of A*s (Ind.) | -0.874 | -0.684 | -1.383 | -1.001 |
| | (0.285)$^{***}$ | (0.192)$^{***}$ | (0.294)$^{***}$ | (0.503)$^{*}$ |
| Norrington Score / 10 | 138.105 | 11.166 | 108.495 | 285.789 |
| | (59.907)$^{**}$ | (33.449) | (66.867) | (158.560)$^{*}$ |
| $N$ | 162 | 162 | 162 | 162 |
| $R^2$ | 0.177 | 0.217 | 0.195 | 0.243 |

*Notes:* Analysis excludes individuals for whom school IDACI is unknown. Policy Off in 2005 and 2006; Policy On in 2008, 2009 and 2010. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate staistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 20: School IDACI of interviewees - changes at the mean, lower quartile, median and upper quartile of colleges' distributions: difference in differences estimates

| | (1) Mean | (2) Q25 | (3) Median | (4) Q75 |
|---|---|---|---|---|
| Constant ($\alpha$) | 0.359 | 6.164 | 4.558 | -25.371 |
| | (11.886) | (5.506) | (13.058) | (30.838) |
| Treated ($\beta$) | -0.249 | -0.155 | 0.072 | 0.766 |
| | (0.386) | (0.288) | (0.475) | (0.778) |
| Policy On ($\gamma$) | 0.260 | 0.410 | 0.392 | 0.288 |
| | (0.503) | (0.281) | (0.632) | (1.247) |
| Treated*Policy On ($\delta$) | 0.532 | 0.005 | 0.174 | 0.375 |
| | (0.431) | (0.319) | (0.421) | (0.927) |
| Mean No. of GCSEs (State) | 0.077 | -0.774 | -1.492 | 3.110 |
| | (1.253) | (0.445)$^{*}$ | (1.409) | (3.241) |
| Mean No. of GCSEs (Ind.) | 0.985 | 0.854 | 1.878 | 0.657 |
| | (0.677) | (0.370)$^{**}$ | (0.722)$^{**}$ | (1.496) |
| Mean No. of A*s (State) | -0.225 | 0.140 | 0.437 | -1.796 |
| | (0.287) | (0.136) | (0.317) | (1.101) |
| Mean No. of A*s (Ind.) | -0.748 | -0.487 | -1.282 | -1.152 |
| | (0.269)$^{***}$ | (0.233)$^{**}$ | (0.307)$^{***}$ | (0.733) |
| Norrington Score / 10 | 120.216 | -3.341 | 102.596 | 331.748 |
| | (68.653)$^{*}$ | (34.824) | (63.018) | (161.009)$^{**}$ |
| $N$ | 162 | 162 | 162 | 162 |
| $R^2$ | 0.096 | 0.148 | 0.193 | 0.160 |

*Notes:* Analysis excludes individuals for whom school IDACI is unknown. Policy Off in 2005 and 2006; Policy On in 2008, 2009 and 2010. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate staistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 21: School IDACI of applicants offered a place - changes at the mean, lower quartile, median and upper quartile of colleges' distributions: difference in differences estimates

| | (1) Mean | (2) Q25 | (3) Median | (4) Q75 |
|---|---|---|---|---|
| Constant ($\alpha$) | 9.266 | 7.591 | 0.264 | 22.195 |
| | (11.152) | (4.368)$^{*}$ | (13.478) | (21.765) |
| Treated ($\beta$) | 0.597 | 0.286 | 0.914 | 2.290 |
| | (0.807) | (0.425) | (0.778) | (1.627) |
| Policy On ($\gamma$) | 0.224 | 0.438 | 0.197 | 0.746 |
| | (0.549) | (0.234)$^{*}$ | (0.669) | (1.148) |
| Treated*Policy On ($\delta$) | -0.493 | -0.304 | -0.943 | -1.466 |
| | (0.890) | (0.403) | (0.844) | (1.919) |
| Mean No. of GCSEs (State) | -0.196 | -0.223 | 0.353 | -0.756 |
| | (1.206) | (0.404) | (1.210) | (2.484) |
| Mean No. of GCSEs (Ind.) | 0.921 | 0.080 | 0.491 | 1.249 |
| | (0.516)$^{*}$ | (0.303) | (0.642) | (1.278) |
| Mean No. of A*s (State) | 0.428 | -0.210 | 0.459 | 0.303 |
| | (0.467) | (0.191) | (0.474) | (0.890) |
| Mean No. of A*s (Ind.) | -0.627 | -0.179 | -1.011 | -0.866 |
| | (0.460) | (0.238) | (0.414)$^{**}$ | (0.927) |
| Norrington Score / 10 | -35.798 | 7.594 | 54.127 | -85.952 |
| | (61.934) | (38.353) | (74.826) | (113.906) |
| $N$ | 114 | 114 | 114 | 114 |
| $R^2$ | 0.051 | 0.061 | 0.085 | 0.046 |

*Notes:* Analysis excludes individuals for whom school IDACI is unknown. Policy Off in 2005 and 2006; Policy On in 2008, 2009 and 2010. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate staistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

children in the average area living in income deprivation per 1000 children. Examining different points of the distribution adds little additional information, since all the estimates are statistically insignificant and show no obvious pattern.

Turning to those called to interview, the results for the mean again concord with those we might expect from the earlier analysis by school type. Table 20 shows no statistically significant difference in the mean IDACI, although the estimate is again positive. Estimates for different points of the distribution are again statistically insignificant from one another or zero, but show some suggestion that the effect is larger in the areas with higher income deprivation (although none are as large as the estimate at the mean).

Finally, considering changes in the mean school-level IDACI of those who get an offer (Table 21) shows somewhat larger absolute estimates than analysis of the interviewees. However, it is worth noting than, unlike at earlier stages and in the analysis of the proportions from state school, the estimates are negative. None of the estimates are statistically significant, so we can have little confidence in this finding, especially as it is inconsistent with most of the analysis.

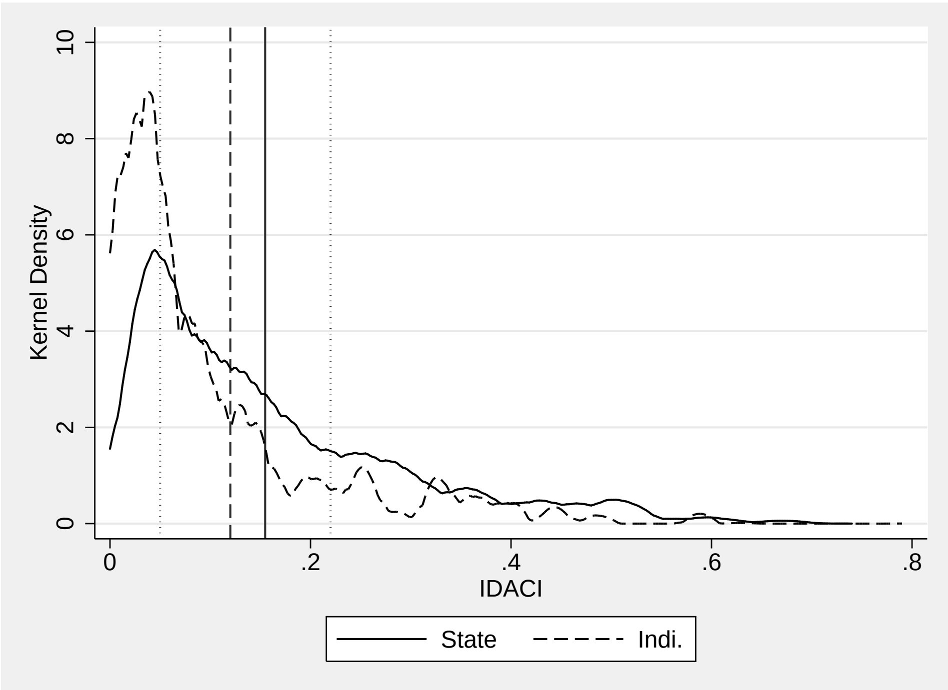
### A.1 Within state school variation

While the above analysis includes all applicants, I now restrict my attention to changes in the distribution of the school-level IDACI just within state school applicants. There is more than one reason for doing this. First, the vast majority of the population attend state schools and the average socioeconomic status of young people attending these schools varies significantly. As such, it would be possible for there to be large changes in the socioeconomic status of applicants, interviewees and those offered a place without observing any changes in variables relating to school type. This analysis assesses whether this is indeed the case.

The second reason is that we might be more concerned about the relevance of the school-area IDACI in the case of independent schools: young people who go to such schools often travel further to attend, particularly as they are far more likely to offer boarding provision. As such, excluding individuals from independent schools may give a more reliable idea about changes in individual-level socioeconomic status using school-level data.

The mean school-level IDACI of applicants from state schools (15%) is higher than that from independent schools (10%). We see the same when considering the median applicant in each case, with IDACI of 12% for the median state school applicant and of 5% for the median independent school applicant. The overall difference in the two distributions is shown by plots of the kernel density of the IDACI for independent and state school applicants in Figure 7.

Figure 7: Kernel density distribution of IDACI by school type




**Notes:** Solid vertical line indicates mean, dashed vertical line indicates median, and dotted vertical lines indicate upper and lower quartiles for state school applicants. Excludes individuals for whom school IDACI was not recorded.

The design of the results tables is the same as those earlier in this section. I report the analyses for each stage of the admissions process in Tables 22, 23 and 24. Since we are only considering those from state school, I only control for the average GCSE performance of state school applicants and college's Norrington score, not the mean performance of independent school applicants.

When it comes to state school applicants, the results for the mean again concord with findings from the analysis in Section 8. We see from Table 22 very little estimated effect on the mean area deprivation level of applicants' schools, although the estimate is positive. Likewise with Table 23 for the mean school-level IDACI among interviewees. In neither case does analysing the quantiles provide any obvious addition to the narrative: in all cases the difference in differences estimates are not statistically significant from either zero or each other.

Finally, I consider the changes in the school-level IDACI of those state school applicants who

Table 22: School IDACI of state school applicants - changes at the mean, lower quartile, median and upper quartile of colleges' distributions: difference in differences estimates

| | (1) Mean | (2) Q25 | (3) Median | (4) Q75 |
|---|---|---|---|---|
| Constant ($\alpha$) | 21.910 | 6.471 | 17.106 | 42.621 |
| | (16.160) | (8.822) | (16.748) | (32.004) |
| Treated ($\beta$) | -0.251 | 0.318 | -0.157 | 0.025 |
| | (0.458) | (0.299) | (0.506) | (0.996) |
| Policy On ($\gamma$) | 0.132 | -0.088 | 0.033 | 0.637 |
| | (0.588) | (0.351) | (0.713) | (1.266) |
| Treated*Policy On ($\delta$) | 0.156 | 0.032 | 0.328 | -0.016 |
| | (0.686) | (0.342) | (0.692) | (1.380) |
| Mean No. of GCSEs (State) | -0.105 | -0.005 | -0.424 | -0.989 |
| | (1.502) | (0.869) | (1.640) | (3.087) |
| Mean No. of A*s (State) | 0.332 | 0.320 | 0.385 | 0.202 |
| | (0.770) | (0.163)$^{*}$ | (0.560) | (1.300) |
| Norrington Score / 10 | -106.586 | -43.790 | -49.398 | -175.889 |
| | (105.165) | (39.161) | (82.454) | (180.250) |
| $N$ | 162 | 162 | 162 | 162 |
| $R^2$ | 0.065 | 0.043 | 0.042 | 0.063 |

*Notes:* Analysis excludes individuals for whom school IDACI is unknown. Policy Off in 2005 and 2006; Policy On in 2008, 2009 and 2010. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate staistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 23: School IDACI of state school interviewees - changes at the mean, lower quartile, median and upper quartile of colleges' distributions: difference in differences estimates

| | (1) Mean | (2) Q25 | (3) Median | (4) Q75 |
|---|---|---|---|---|
| Constant ($\alpha$) | 18.853 | 5.060 | 11.968 | 23.917 |
| | (16.708) | (12.625) | (23.914) | (26.112) |
| Treated ($\beta$) | -0.208 | 0.318 | 0.154 | 0.843 |
| | (0.474) | (0.376) | (0.629) | (1.113) |
| Policy On ($\gamma$) | -0.606 | -0.447 | -0.566 | -0.413 |
| | (0.827) | (0.570) | (1.168) | (1.226) |
| Treated*Policy On ($\delta$) | 0.088 | 0.104 | -0.243 | -0.883 |
| | (0.781) | (0.506) | (0.846) | (1.788) |
| Mean No. of GCSEs (State) | -0.128 | 0.317 | 0.207 | -0.012 |
| | (1.503) | (1.131) | (2.249) | (2.217) |
| Mean No. of A*s (State) | 0.621 | 0.407 | 0.561 | 0.093 |
| | (0.551) | (0.168)$^{**}$ | (0.496) | (1.089) |
| Norrington Score / 10 | -88.883 | -81.945 | -88.896 | -47.000 |
| | (104.392) | (39.732)$^{**}$ | (84.930) | (222.590) |
| $N$ | 162 | 162 | 162 | 162 |
| $R^2$ | 0.038 | 0.057 | 0.021 | 0.008 |

*Notes:* Analysis excludes individuals for whom school IDACI is unknown. Policy Off in 2005 and 2006; Policy On in 2008, 2009 and 2010. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate staistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 24: School IDACI of state school applicants offered a place - changes at the mean, lower quartile, median and upper quartile of colleges' distributions: difference in differences estimates

| | (1) Mean | (2) Q25 | (3) Median | (4) Q75 |
|---|---|---|---|---|
| Constant ($\alpha$) | 40.251 | 13.050 | 22.562 | 47.482 |
| | (16.971)** | (7.928) | (15.437) | (32.317) |
| Treated ($\beta$) | 0.507 | 0.672 | 0.286 | 2.469 |
| | (0.929) | (0.502) | (1.040) | (1.886) |
| Policy On ($\gamma$) | 0.305 | 0.103 | -0.171 | 0.329 |
| | (0.852) | (0.425) | (0.768) | (1.539) |
| Treated*Policy On ($\delta$) | -0.908 | -0.692 | -1.104 | -2.362 |
| | (1.077) | (0.662) | (1.027) | (2.226) |
| Mean No. of GCSEs (State) | -1.666 | -0.422 | -0.589 | -0.990 |
| | (1.715) | (0.814) | (1.582) | (3.146) |
| Mean No. of A*s (State) | 0.751 | 0.040 | 0.074 | 0.753 |
| | (0.518) | (0.291) | (0.536) | (1.048) |
| Norrington Score / 10 | -199.970 | -58.965 | -87.129 | -326.103 |
| | (105.383)* | (64.206) | (92.494) | (185.568)* |
| $N$ | 116 | 116 | 116 | 116 |
| $R^2$ | 0.063 | 0.050 | 0.043 | 0.057 |

*Notes:* Analysis excludes individuals for whom school IDACI is unknown. Policy Off in 2005 and 2006; Policy On in 2008, 2009 and 2010. Standard errors, clustered by college-subject group combination, in parentheses. Stars indicate staistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

get an offer (Table 24). As with the analysis of all those offered a place, the change in mean IDACI of those from state schools offered a place is estimated to be negative. However, this time the estimate is rather larger, but still far from statistical significance.

## A.2 Discussion

Analysis considering changes at different quantiles is more difficult to interpret a single estimate of changes in means. However, its results have the potential to provide more information on the nature of the impact.

In this analysis, while the point estimates at different quantiles do vary from one another and from the estimated changes in means, these differences are never statistically significant from zero or each other. Nevertheless, that we see some variation is suggestive of differing impacts across the deprivation distribution. Furthermore, there is little sign of a consistent pattern towards one end of the distribution or the other.

Nevertheless, the point estimates we see tend to back up the story of very little socioeconomic change resulting from the introduction of the TSA, as seen in the main analysis.