# The Impact of Standardised Testing on Later High Stakes Test Outcomes

Joseph Regan-Stansfield

The Department of Economics
Lancaster University Management School
Lancaster LA1 4YX
UK

LUMS home page: http://www.lancaster.ac.uk/lums/

# The Impact of Standardised Testing on Later High Stakes Test Outcomes

**Joseph Regan-Stansfield[1]**
**Department of Economics, Lancaster University**

**March 2017**

### Abstract

Standardised tests are a common, yet contentious, feature of many countries' schooling systems. In May 2010, over one-quarter of English primary schools boycotted that year's mandatory age eleven standardised tests (colloquially known as SATs tests). This paper investigates the plausibly causal effect of participation in standardised testing on later end-of-schooling qualification (GCSE) attainment. After controlling for non-random boycott participation, and relying on a selection-on-observables argument, evidence is found of a statistically significant negative effect of boycott participation on various measures of GCSE attainment. Amongst other findings, pupils are estimated to be 0.7 per-cent less likely to achieve five or more GCSEs at grades A* to C due to not sitting their age eleven SATs tests. Thus, evidence is found that individual pupils' subsequent attainment benefits from their own prior participation in standardised tests.

**Keywords:** Education Policy, Standardised Testing, Secondary Education

**JEL codes:** I21, I28, J24

---

1

# 1    Introduction

A near ubiquitous feature of schooling systems in developed nations today is mandatory standardised tests. These tests, frequently sat by pupils at critical stages of the schooling process, are often an integral component of the school accountability system. In England, for example, the aggregated results from national curriculum assessment tests (colloquially known as SATs tests) are the headline measure of school performance reported in primary school league tables.

The nascent literature on school accountability systems generally suggests that they have a positive effect on pupil performance (Figlio and Loeb 2011). Using English and Welsh data, Burgess *et al.* (2013) finds evidence that increased school accountability has a significant positive causal effect on GCSE attainment. Findings such as this suggest that there is a social benefit from pupils participating in standardised tests. The information provided by participating pupils allows regulators to identify under-performing schools, while simultaneously standardised testing incentivises all schools to maximise their teaching performance. Both mechanisms subsequently benefit all pupils even those who may not sit the standardised test.

However, it is largely unknown whether there is any private benefit to pupils from standardised testing. That is, does an individual pupil's subsequent attainment improve if they sit the standardised test? Or equivalently, is a pupil's attainment unaffected if they do not participate in standardised testing, given that they can free-ride from their peers' participation in such tests?

This paper attempts to quantify the causal effect of participation in standardised assessments for individual pupils. To this end, a widespread headteacher-led boycott of one series of mandatory standardised tests in England is exploited as a natural experiment. The boycott prevented over one-quarter of the affected cohort from participating in a series of mandatory age eleven SATs tests. I

compare the differences in GCSE (England's school leavers' qualification) outcomes for the affected cohort between those who did and did not participate in age eleven SATs.

After controlling for school-level and pupil-level confounding variables, and depending on a selection-on-observables assumption, I find evidence of a negative effect on GCSE attainment for pupils who attended primary schools that boycotted the age eleven SATs tests. In my preferred specification, pupils who attended boycotting primary schools are on average 0.76 per-cent less likely to obtain five or more GCSEs (or equivalent qualifications). This translates into around 4,000 pupils failing to meet the *de-facto* minimum standard to pursue the academic track at further education. Not participating in KS2 SATs tests is estimated to reduce the average pupil's point score by 0.16 and 0.29 points in GCSE English and maths respectively. This is equivalent to 2.68 per-cent of pupils dropping one grade at GCSE English, and 4.86 per-cent of pupils dropping a GCSE maths grade. Therefore, I find evidence that pupils experience a private benefit to participation in mandatory standardised tests.

The boycott was led by two teachers' unions which balloted headteachers and their deputies over the decision to take this form of industrial action. Ballot results were announced less than one month prior to the tests being administered. It is likely that teachers prepared their pupils for KS2 SATs tests in a comparable fashion – for most of the school-year – irrespective of whether their school ultimately boycotted SATs tests. Therefore, the estimated boycott participation effect solely reflects the consequence of not having a valid KS2 SATs test result. It does not capture the effects of preparation for SATs tests on later exam outcomes.

Standardised testing has increasingly been criticised by teachers, parents and academics. The nature of the criticism is diverse. One concern is that by introducing standardised testing at early stages of schooling, children are being subjected to "test anxiety" or "exam stress" from an

increasingly young age (Connor 2001, 2003). Another prominent criticism of standardised testing is that teachers become incentivised to focus on preparing their pupils for tests at the detriment of teaching a broad and balanced curriculum (so called "teaching to the test"). A theoretical literature has highlighted this issue (Koretz 2002, Lazear 2006). In a similar vein, Neal (2010) highlights that given variation in pupil ability and diminishing marginal returns to teacher effort on pupil attainment, it is inefficient to encourage teachers to aim to prepare all students to reach a common level of attainment in standardised tests. Furthermore, in England, standardised tests have been identified as a source of demotivation amongst teachers, and a contributing factor of England's teacher retention problem (Day and Smethem 2009). Whilst these issues are valid reasons to doubt the wisdom of mandatory standardised testing, this paper suggests that by the end of compulsory schooling, pupils' attainment will have benefited from their own prior participation in standardised tests.

Since within school tracking by ability is not uncommon in English secondary schools, the class allocation process may represent one mechanism through which standardised test participation effects future attainment. KS2 SATs results are a cheap, usually readily available, and – in theory – reliable measure of pupil ability. The unavailability of these results for a significant proportion of the affected cohort may have increased the likelihood that the secondary school misjudged the ability of an incoming pupil, and subsequently assigned him/her to an unintended class. This may have caused under attainment relative to the counterfactual of assignment to the correct ability group class. However, this argument presumes tracking positively affects pupil attainment; the economics of education literature is yet to come to a definite consensus on this issue (Betts 2011).

KS2 SATs tests are high stakes from the perspective of schools and teachers, and are conducted under strict formal exam conditions. They represent the first and last opportunity for pupils to

experience such an examination environment before they sit their GCSE examinations. Participation in KS2 SATs plausibly improves future attainment through enhancing the pupil's familiarity with the process of sitting high stakes exams (including preparation for, and completion of, the test).

A third potential mechanism is SATs test results provide a useful source of feedback to pupils, parents and teachers. Parental and pupil inputs into the education production function are likely to be endogenous to this information. If a pupil (or their parents) underestimate their relative attainment, then SATs test feedback can be used to recalibrate this estimate and consequentially their effort level. Test results may also increase the efficiency of teachers' lesson planning. Indeed, Dobbie and Fryer Jr (2013) and Fryer Jr (2014) both find evidence of a positive attainment return from data-driven instruction.

## 2    Institutional background

The precursor to mandatory standardised assessment of pupils in English schools was the establishment in law of the Education Reform Act in 1988. The Act harmonised the curriculum and organisational structure of schools across England. The legislation introduced the national curriculum which all state schools were expected to deliver, and defined four key stages (KS) of schooling: five to seven years old (KS1); eight to eleven years old (KS2); twelve to fourteen years old (KS3), and fifteen to sixteen years old (KS4).

Prior to the Act, pupils in English schools were formally assessed only at the conclusion of secondary schooling (also the end of KS4) when they sat examinations in national recognised O-Level or CSE qualifications.[2] However, in order to support the national curriculum, the Act

---

[2] In 1988, O-Level and CSE qualifications were discontinued in favour of GCSE (General Certificate of Secondary Education) qualifications.

specified that pupils should be assessed at the end of each key stage "for the purpose of ascertaining what they have achieved in relation to the attainment targets for that stage" (Education Reform Act, 1, 2 (2)).

The Task Group on Assessment and Testing (TGAT) was responsible for developing the new assessment system. They determined that the system had to satisfy several distinct purposes. The system should: provide information on the achievement of pupils; enable teachers to plan the next stage(s) for pupils; provide information on the aggregated achievement of pupils (to evaluate the functioning of schools and teachers); and provide information to parents to inform school choice decisions (Whetton 2009).

KS1 assessment was introduced in 1991, while KS2 and KS3 assessment followed in 1994. At that time, national curriculum assessment (SATs) consisted of tests and teacher assessments (which had notionally equal status) at all key stages. The tests are externally marked, and are subject to stringent procedures for maintaining standards.

The assessments measure attainment in terms of national curriculum levels, which range from one to eight. Pupils are expected to be working at level two at the end of KS1. Pupils should make two levels of progress between each key stage, meaning that pupils should achieve levels four and six at the end of KS2 and KS3 respectively. This information is summarised in Table 1. Pupils are assessed in the core subjects of English, mathematics and science. English assessment consists of an overall national curriculum level, and separate levels for reading, writing, speaking and listening, and spelling, punctuation and grammar depending on the key stage and policy at the time of assessment.

The arrangements for SATs were largely unchanged until 2005, when the government reformed KS1 assessment and dropped tests in favour of more detailed teacher assessments. Testing at KS3 met the same fate in 2008 when the government concluded that parents obtained the same information from GCSE test results as KS3 test results. KS2 testing continued, although from 2009 only a subset of schools is required to administer science tests for the purpose of monitoring national standards.

## 2.1 The Boycott

In May 2010, 3,942 primary schools participated in a boycott of KS2 SATs tests. This represents 26.79 per-cent of the 14,716 mainstream primary schools in England that were expected to administer the KS2 SATs tests that year. Thus, 27.47 per-cent of the Year 6 cohort from 2009/10 did not have a valid SATs test result in any of the English or maths tests. In comparison, fewer than 0.5 per-cent of 2008/09's Year 6 cohort did not have a valid test result for English and maths.

The boycott was organised by the two biggest teachers' unions in England, the National Union of Teachers (NUT) and the National Association of Head Teachers (NAHT). The unions jointly balloted members who were head teachers or deputy head teachers of primary schools.

61 per-cent of NAHT voters and 74.9 per-cent of NUT voters voted in favour of the boycott. Turnout was 49.7 per-cent for NAHT members against 33.8 per-cent for members of the NUT. Approximately, 25,000 head teachers or deputy head teachers were eligible to participate in the ballot which represents the majority of primary school teachers in leadership positions.

The result of the ballot was announced on 16th April 2010 while the tests were due to be sat during the week commencing 10th May 2010. Given this short window between the decision to boycott and the commencement of the tests, it is unlikely that the experience of pupils in schools that

participated in the boycott was very different over the school-year to pupils in the schools that did not boycott. The pupils in the boycotting primary schools still likely prepared for SATs tests for much of the school-year. Additionally, the boycott did not affect the secondary school that the pupil would subsequently attend; pupils were allocated to secondary schools in March.

## 3    Data

This paper uses an extract from the Department for Education's National Pupil Database (NPD), a well-established collection of administrative datasets covering all of England's state schools and their pupils. For the main analysis, I use two waves of the Spring School Census, from the 2009/10 and 2014/15 school-years to link pupils to the schools they attend in the final years of their primary and secondary schooling. The Spring School Census also contains rich demographic data such as gender, ethnicity, first or native language; special education needs (SEN) status, and month and year of birth which I include in my models as control variables. The best available indicator of the pupil's socioeconomic circumstances is a pupil's history of eligibility for free school meals (FSM).

To control for observable differences between primary schools that participated in the boycott and those that did not, I utilise primary school level averages of several pupil characteristics. I additionally include measures of the primary school level average attainment of the cohort which immediately preceded the boycott affected cohort, as well as the average prior (KS1, age seven) attainment of the boycott affected cohort.

The Spring School Census data is linked to attainment datasets. I report models for three outcome variables, each of which are measures of GCSE exam attainment at the end of KS4/secondary school. The first outcome is the point score achieved by the pupil in GCSE English, the second is the point score achieved in GCSE maths. GCSEs are awarded with letter grades A*, A, …, G. In

8

the point score measure, A* is coded as 58 points and G is coded as sixteen points meaning that one grade is equivalent to six points. This measure implicitly assumes that moving from grade A to A* is comparable to moving from grade G to F, and so on. I opt to study the effect of the boycott on these two GCSE qualifications since (almost) all pupils study maths and English until the age of sixteen, and because these are the two broad subjects which are tested by the KS2 SATs tests.

The third outcome variable measures attainment at GCSE level more broadly. It is a binary variable equal to one if the pupil achieved five or more GCSEs (or equivalents) at grades A* to C, or zero otherwise. This is the standard all pupils are expected to achieve at the end of their compulsory schooling. The proportion of pupils who achieve this threshold is an important measure used by parents to infer the "quality" of secondary schools.

Table 2 presents the means of primary school level averages of several pupil characteristics by boycott participation. The mean school level average prior (KS1) attainment of the affected cohort and the average KS2 attainment of the cohort immediately preceding the affected cohort are also included in the table. The third column contains the difference between the means of boycotter and non-boycotter primaries, and denotes whether the means are statistically different at the one, five or ten per-cent significance levels.

The difference in the mean of a covariate between the boycotter and non-boycotter primary schools is regularly statistically significant at the one per-cent level. Schools that boycotted SATs performed worse in KS2 English and maths tests in the year prior to the boycott, than those that did not boycott. In addition, the average prior attainment (in age seven KS1 teacher assessments) of the affected cohort is lower in the boycotting schools than non-boycotters. Despite performing less well than non-boycotters in KS2 and KS1 assessments, boycotter primary schools are more

9

likely to be rated good by OFSTED (the schools' regulator) and less likely to be rated inadequate compared to non-boycotters.

In terms of student demographics, boycotter primary schools tend to have more pupils and have proportionally more pupils eligible for, and taking, free school meals. They also have proportionally fewer white British pupils, and pupils whose first language is English.

Boycott participation is non-random; it was at every school's discretion whether to participate. As such, it is unsurprising that the distributions of these covariates are not balanced across both groups of schools. Two factors drive the difference in covariate distributions between the groups. First, poorer performing schools have a greater incentive to disrupt the school accountability system, specifically school league tables; while schools with a record of good performance in KS2 SATs tests risk greater reputational harm from participation than under-performing schools. Secondly, urban schools were more likely to participate in the boycott than non-urban schools. The NAHT speculates that this is because it is easier for urban schools to club together and jointly decide to boycott, whereas rural headteachers might be less well connected to fellow headteachers.

Figures 1a and 1b show, respectively, the distribution of the school-level average KS2 English and maths point score of the cohort immediately preceding the boycott affected cohort, split by boycott participation. Like Table 2, both figures suggest that the schools that boycotted had previously worse performance in KS2 SATs tests relative to those that did not boycott.

## 4    Methodology

I estimate the following equation using OLS,

$$y_{isp} = \alpha_s + \beta_1 Boycotter_p + \beta_2' X_i + \beta_3' X_p + \varepsilon_{isp} \tag{1}$$

where $i$, $s$, and $p$ are pupil, secondary school and primary school identifiers respectively. $y_{isp}$ refers to a measure of pupil GCSE attainment, $\alpha_s$ is a secondary school fixed effect. It is not possible to include a primary school fixed effect since it would be perfectly collinear with $Bocycotter_p$, which is equal to one if primary school $p$ participated in the SATs Boycott, and zero otherwise. $\beta_1$ is the estimate of the association between boycott participation and KS4 attainment. Vectors $X_i$ and $X_p$ refer to pupil and primary school level control variables respectively. The pupil's prior attainment, as measured by their KS1 teacher assessment, is included in $X_i$. $\varepsilon_{isp}$ denotes the error term.

Recognising the clustered nature of pupil-level data within primary school and secondary school year-groups, I estimate heteroscedasticity robust two-way clustered standard errors. There are approximately 13,000 clusters in the primary school dimension, and approximately 4,000 clusters in the secondary school dimension.

Assuming selection-on-observables, $\hat{\beta}_1$ represents the average treatment effect on treated (ATET). If, however, conditional on the control variables, treatment is not as good as randomly assigned then $\hat{\beta}_1$ is biased. Since schools self-selected into treatment, it is necessary to include primary school level covariates to control for the underlying differences between treatment (boycotter) and control (non-boycotter) primary schools. Without these controls the bias is likely to be negative; less effective primary schools are a have greater incentive to participate in the boycott. In later sections, I search for evidence that the selection-on-observables assumption does not hold. No such evidence is found which should increase confidence in this assumption.

### 4.1 Moving towards covariate balance with matching

Table 2 shows that the means of school level covariates are statistically different between schools that did and did not participate in the boycott at the one per-cent significance level. The imbalances in pre-treatment confounding covariate distributions between these two groups of primary schools must be controlled for. This can be achieved through modelling the effect of these confounding covariates in the OLS specification. As an additional step, however, I also combine regression with a matching procedure: coarsened exact matching (CEM).

CEM is a non-parametric data pre-processing method designed to enhance the empirical covariate distribution balance between treatment (boycotter primary schools) and control (non-boycotter primary schools) groups (Iacus *et al.* 2012). Exact one-to-one matching involves matching treated observations to control observations which have common covariate values. Control and treatment observations which do not match to observations from the other group are excluded from the analysis, and the treatment effect is estimated only using matched treatment and control observations.

Exact matching, however, is often not a feasible solution. A large set of continuously valued control variables will severely reduce the likelihood of a match between a treated observation and a control observation to the extent that matches will become almost non-existent. In CEM, covariates are temporarily coarsened or aggregated into non-overlapping bins, and a stratum is defined for each permutation of the coarsened covariates. Observations are then assigned to a stratum based on the values of their covariates. Observations which belong to a stratum containing at least one control observation and one treatment observation are "matched". Observations belonging to a stratum which do not contain both treatment and control observations are not "matched" meaning they are dropped from the sample.

12

The empirical distribution of the covariates between the treated and control groups within the processed (or "matched") sample should be more balanced compared to the distributions from the unprocessed (or "full") sample. I apply the regression model presented in the preceding section to the matched sample.

Since treatment assignment is at the school level, I perform CEM on school level data, meaning that I attempt to match primary schools that participated in the boycott to primary schools that did not participate in the boycott. The regression model is then estimated on pupil level data from pupils who attended the primary schools that are in the matched sample of schools. I match schools on the basis of: the average KS2 attainment of the pre-boycott cohort (2008/09); the average prior (KS1) attainment of the boycott cohort (2009/10); the number of pupils; the number of FTE qualified teachers; and the percentage of pupils eligible for FSM eligibility. These variables are used as the basis for matching for two reasons. First, they address different dimensions of school level covariates (attainment of a past cohort, prior attainment of the current cohort, school size, school resource and average pupil socioeconomic circumstance). Second, the distribution of these covariates exhibited the greatest imbalance between boycotter and non-boycotter primary schools in the full sample (as defined by the univariate L1 distance).

CEM is adopted since it has several attractive properties compared to other matching procedures. CEM belongs to a class of matching methods known as "Monotonic Imbalance Bounding" (Iacus *et al.* 2011). The maximum degree of imbalance is known *ex ante*: the greater the coarsening of the covariates, the greater the imbalance in the covariate distribution permissible. Secondly, the degree of coarsening for one variable does not affect the imbalance bound for any other variable. In other words, it is possible to increase the balance on one covariate (through lessened coarsening) without sacrificing balance on the other covariates.

13

## 5    Full sample results

Table 3 features estimates from five specifications of an OLS model for GCSE English point score. The coefficient point estimate for the boycott participation indicator variable is found in the first row of the table. Column 1 presents the coefficient estimate from a naïve model without any control variables or secondary school fixed effects. The coefficient estimate is -0.582 which is statistically different from zero at the one per-cent significance level. My expectation is that this estimate is biased downwards, and that the coefficient point estimate should become less negative as control variables are added to the model. In Column 2 secondary school fixed effects are added to the model, and, as anticipated, the boycott participation coefficient estimate increases to -0.341. In Column 3 I add primary school level controls. These control coefficient estimates are estimated with high precision, only two coefficient estimates are not statistically different from zero at the five per-cent significance level. The boycott participation effect estimate halves to -0.160, and is again statistically different from zero. In Column 4, I add pupil level controls. All pupil level control variable coefficients are precisely estimated at the one per-cent significance level. In this specification, the boycott participation coefficient point estimate is statistically different from zero at -0.1155.

The last column (the preferred specification) adds a measure of the pupil's prior attainment. The preferred boycott participation indicator coefficient estimate of -0.161 is statistically different from zero at the one per-cent level. A 0.161 reduction in GCSE English point score is equivalent to losing 2.68 per-cent of a grade in GCSE English. It is also comparable to 1 in 37 pupils who attended a boycotting primary school losing one grade in GCSE English.

Table 4 shares the format of the preceding table; however, the outcome variable is GCSE maths point score. Column 1 reports the boycott participation coefficient estimate in a model absent of

14

control variables. The point estimate of -1.0404 is highly precise. Once secondary school fixed effects are added in Column 2 the point estimate halves to -0.5377 and remains statistically different from zero at the one per-cent significance level. When primary school controls and pupil level controls (including prior attainment) are added in Column 5, the coefficient estimate of interest is -0.291. This point estimate is precise; the p-value of the null hypothesis that the underlying population coefficient is zero is 0.00001. The effect of boycott participation on GCSE maths point score is therefore estimated to be eighty per-cent larger than the corresponding effect on GCSE English point score. This is consistent with the usual pattern within the education economics literature of larger estimated treatment effects on maths outcomes than on English or reading outcomes. The estimated 0.291 decrease in GCSE maths point score is tantamount to losing 4.85 per-cent of a GCSE maths grade, or 1 in 21 pupils educated at boycotting primary schools losing one grade in GCSE maths due to the boycott.

The outcome variable modelled in Table 5 is a binary variable equal to one if the pupil achieved five or more GCSEs (or equivalents) at grades A* to C, or zero otherwise. The coefficient estimate for boycott participation is once again estimated with high precision; all estimates are statistically different from zero at the one per-cent significance level. The point estimate from the model without control variables is -0.029. The point estimate increases as secondary school fixed effects, and primary school and pupil level control variables are included in the model.

In the preferred specification (Column 5), the point estimate is -0.0076. This suggests that pupils who attended a primary school which failed to fully administer their KS2 SATs tests are 0.76 percentage points less likely to achieve the minimum expected GCSE performance at age 16 than their peers who participated in the KS2 SATs tests. In the sample 66.66 per-cent of pupils achieve

15

the threshold, therefore the boycott participation effect is equivalent to a one per-cent decline in the pass rate.

In Table 6, I test for the presence of heterogeneous treatment effects at both the pupil level and the secondary school level. Columns 1 to 3 contain estimates from a heterogeneous treatment effects model for GCSE English point score, GCSE maths point score is the outcome variable in columns 4 to 6, while the outcome in the remaining columns is the five or more GCSEs (or equivalents) at grades A* to C indicator. Column 1 indicates that the effect on GCSE English point score of attending a primary school that participated in the boycott does not vary between pupils with and without special educational needs. Colum 2 similarly reports that the effect on GCSE English point score does not vary with the pupil's free school meal eligibility.

Column 3 investigates whether the effect of attending a boycott primary school varies by the "quality" of the secondary school attended by the pupil – as measured by the secondary schools OFSTED rating. I group schools that have an outstanding or good rating together; 67 per-cent of secondary schools in the sample fall into this category. Inadequate and "requires improvement" schools are also grouped together (this group is the omitted category). The coefficient estimate on the interaction between the boycott participation indicator and the outstanding/good OFSTED rating indicator is not statistically different from zero at conventional significant levels. Furthermore, the boycott participation indicator is no longer statistically different from zero at even modest significance levels. There is no evidence that the boycott participation effect on GCSE English point score varies by the "quality" of the secondary school attended by a pupil.

Column 4 suggests that the estimated boycott participation effect on GCSE maths point score is homogenous with respect to SEN status. Column 5, however, shows that the estimated boycott effect does vary with FSM eligibility. The boycott participation effect is estimated at -0.2523

points for FSM eligible pupils. The effect on FSM ineligible pupils is considerably smaller at approximately -0.046 points, and is statistically significant at the five per-cent significance level. Column 6 presents estimates from a GCSE maths point score model which allows the boycott participation effect to vary by the OFSTED rating of the pupil's secondary school. Unlike the corresponding model for GCSE English point score, the estimate of the boycott participation effect is statistically significant at the one per-cent significance level. This effect is invariant across pupils who attend high and low "quality" secondary schools.

Columns 7 to 9 investigate the presence of a heterogenous treatment effect on the likelihood of achieving five or more GCSEs at grades A* to C. Columns 7 and 8 present surprising results, the boycott participation effect is estimated to be -0.07 per-cent for non-SEN pupils as well as FSM ineligible pupils. But, the boycott participation effect is estimated to be positive for SEN pupils (+0.5 per-cent) and FSM eligible pupils (+0.6 per-cent); all treatment effects are statistically different from zero at the one per-cent significance level. The positive treatment effect for SEN pupils and FSM eligible pupils is puzzling, I speculate that this positive effect may be driven by qualification choice. SEN and FSM eligible pupils are more likely to be towards the lower end of the attainment distribution, and, therefore, they are more likely to study vocational orientated qualifications which might be easier to obtain than the more academically orientated GCSE qualifications. If not participating in SATs harmed the progress of these pupils, then perhaps they become more likely to study vocational qualifications which ultimately increases their likelihood of achieving five or more GCSEs or equivalents at A* to C. Column 9 suggests that the boycott participation effect, precisely estimated as -0.7 per-cent is invariant to the "quality" of the secondary school attended by the pupil.

17

## 5.1    Falsification tests

Identification of the SATs boycott effect relies on a selection-on-observables assumption. This assumption is intrinsically untestable; however, this section presents evidence that enhances its credibility.

If the coefficient estimate on the boycott participation indicator provides an unbiased estimate of the boycott participation effect, then the coefficient estimate derived from a sample of students who were not affected by the boycott should be insignificantly different from zero. If the boycott participation indicator coefficient estimate is significantly different from zero when estimated on an untreated sample, then that estimate is undoubtedly biased, and it is likely that the estimate from the treated sample is subject to a similar bias.

I perform a falsification test by estimating the OLS models on the cohort of students who preceded the treated cohort.[3] The treated cohort was in year group 6 in 2009/10, and year group 11 in 2014/15. The preceding – and untreated – cohort was in year group 6 in 2008/09, and year group 11 in 2013/14. The coefficient on the boycott participation indicator should be insignificantly different from zero when estimated on data for this cohort.

Panel A in Table 7 contains estimates from models for GCSE English point score based on data from the cohort immediately preceding the treated cohort. The boycott participation indicator coefficient estimate is precisely estimated as -0.4892 when the model includes no control variables. However, when primary school-level control variables are included (in columns 3, 4 and 5), the coefficient estimate is a precisely estimated zero effect. Panel B features estimates from GCSE math point score models. Once again, the naïve estimate of the boycott participation indicator

---

[3] Future drafts will contain a falsification test on the cohort immediately succeeding the treated cohort once access to the relevant data has been granted.

coefficient is significantly different from zero at the one per-cent significance level (-0.7776). But when primary school control variables are added to the model, the boycott participation coefficient estimates increase to zero. Finally, Panel C presents estimates from the 5+ GCSE or equivalents at A*-C grade model. The boycotter participation indicator coefficient estimates are precise zeros when primary school-level control variables are included in the model. The models which include primary school control variables (which includes the preferred specification) therefore "pass" the falsification test. Whilst this does not prove that the selection-on-observables assumption is valid, any bias in the estimated effects should have been relatively easy to detect from this exercise.

## 6.     Matched sample results and robustness

Table 8 presents the means of primary school level averages of several pupil characteristics and attainment measures conditional on boycott participation for the restricted sample derived from the coarsened exact matching pre-processing. Unlike the full sample, the difference in the covariate means between the two groups are much less likely to be statistically significant. Relative to the full sample, the covariate distributions are far more balanced across the boycotter and non-boycotter groups. Figures 2a and 2b show the kernel density estimates of the distributions of primary school average KS2 English and maths attainment for the 2008/09 cohort for the two groups of schools. The difference in the distributions for the two groups is much less pronounced for the matched sample than the full sample (see Figures 1a and 1b).

The matching procedure does not resolve the issue of possible selection-on-unobservables. However, since in observable dimensions, the boycotter and non-boycotter schools are more similar in the matched sample than in the full sample, it is fair to assume that the difference between the two groups of schools in unobservable dimensions is also reduced in the matched

sample. Therefore, the internal validity of the estimates from the matched sample is at least as great as that of the full sample's estimates.

The matched sample estimates must have less external validity than the full sample estimates, as only a subset of primary schools (and their pupils) are included in the analysis. Note, in addition to the 7,288 non-boycotter primary schools that are not matched to boycotter primary schools, 456 boycotter primary schools are not matched to non-boycotter primary schools. It would be incorrect to treat the estimates from the matched sample analysis as unambiguously superior to the full sample estimates. For this reason, I principally treat the matched sample analysis as a robustness exercise. If a statistically significant negative effect of boycott participation is not found using the matched sample, in which selection on unobservables is arguably less of a concern, then this may raise concerns over the credibility of the estimated effect in the full sample analysis. If similar estimates arise from the matched sample, then this should increase confidence in the selection-on-observables assumption. Although, for the avoidance of doubt, this assumption remains fundamentally untestable.

Table 9 presents coefficient estimates of the boycott participation indicator under various model specifications applied to the matched sample. Panel A contains estimates of the effect of boycott participation on GCSE English point score. The estimate from the preferred specification (column 5) is -0.197. At the one per-cent significance level, this estimate is statistically different from zero, but not from the corresponding full sample estimate. Panel B features estimates of the boycott participation effect on GCSE maths point score. The preferred estimate is -0.353 which once again is statistically different from zero (but not its full sample counterpart) at the one per-cent significance level. Finally, the estimate of the effect on the likelihood of achieving five or more GCSEs at A* to C is found in Panel C. Pupils are estimated to be 1.3 per-cent less likely to achieve

this threshold because of failing to sit their KS2 SATs test. This estimate is also statistically different from zero at the one per-cent significance level.

As a final exercise, I repeat the falsification test that I conducted on the full sample on the matched sample: I estimate the effect of boycott participation using data on pupils who attended the schools in the matched sample the year prior to the boycott. The results of this exercise are reported in Table 10. Panel A contains the estimates of the effect on GCSE English point score, Panel B contains estimates of the GCSE maths point score effect, and Panel C shows the estimated effect on the likelihood of obtaining five or more GCSEs (grades A* to C). The estimated effect on all outcomes from the preferred specification is not statistically different from zero at conventional significance levels. In fact, only two of the eighteen estimated effects contained in Table 10 are significantly different from zero at the five per-percent significance level.

# 7    Conclusion

This paper considers whether there is a private benefit to pupils from their participation in mandatory standardised tests in terms of their own future end-of-schooling qualification exam performance.

Industrial action by teachers in the form of a widespread boycott of the 2010 KS2 SATs tests is exploited as a natural experiment. I find evidence that pupils who were prevented from participating in the age eleven SATs tests by virtue of the boycott performed less well in their GCSE examinations five years later. Pupils who did not participate in age eleven SATs were 0.7 per-cent less likely to obtain five or more GCSEs (or equivalents) at grades A* to C. 2.85 per-cent of pupils lost one grade in GCSE English as a consequence of the industrial action, while 4.85 per-cent of pupils achieved one grade lower in GCSE maths. Taken together, these findings provide

21

evidence that there is a significant private benefit to pupils from participation in mandatory standardised tests.

The issue of standardised testing is a contentious policy debate internationally. The private costs to pupils of standardised assessment tests, such as test anxiety, and the experience of being taught to the test, are well documented and are often cited by opponents of such tests. Meanwhile there is comparatively less evidence of the benefits of standardised testing. This paper narrows this deficit by providing robust empirical evidence of a private attainment benefit to pupils of sitting such tests.

# References

Betts, J.R., 2011. The Economics of Tracking in Education. *In*: E. Hanushek, S. Machin, and L. Woessmann, eds. *Handbook of the Economics of Education*. 341–381.

Burgess, S., Wilson, D., and Worth, J., 2013. A natural experiment in school accountability: The impact of school performance information on pupil progress. *Journal of Public Economics*, 106, 57–67.

Connor, M.J., 2001. Pupil stress and standard assessment tasks. *Emotional and Behavioural Difficulties*, 6 (2), 103–111.

Connor, M.J., 2003. Pupil stress and standard assessment tasks (SATs) An update. *Emotional and Behavioural Difficulties*, 8 (2), 101–107.

Day, C. and Smethem, L., 2009. The effects of reform: Have teachers really lost their sense of professionalism? *Journal of Educational Change*, 10 (2–3), 141–157.

Dobbie, W. and Fryer Jr, R.G., 2013. Getting Beneath the Veil of Effective Schools: Evidence From New York City. *American Economic Journal: Applied Economics*, 5 (4), 28–60.

Figlio, D. and Loeb, S., 2011. *School Accountability*. 1st ed. Handbook of the Economics of Education. Elsevier B.V.

Fryer Jr, R.G., 2014. Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments. *The Quarterly Journal of Economics*, 129 (3), 1355–1407.

Iacus, S.M., King, G., and Porro, G., 2011. Multivariate Matching Methods That Are Monotonic Imbalance Bounding. *Journal of the American Statistical Association*, 106 (493), 345–361.

Iacus, S.M., King, G., and Porro, G., 2012. Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20 (1), 1–24.

Koretz, D., 2002. Limitations in the Use of Achievement Tests as Measures of Educators' Productivity. *Journal of Human Resources*, 37 (4), 752–777.

Lazear, E., 2006. Speeding, terrorism, and teaching to the test. *The Quarterly Journal of Economics*, 121 (3), 1029–1061.

Neal, D., 2010. Aiming for Efficiency Rather than Proficiency. *Journal of Economic Perspectives*, 24 (3), 119–132.

Whetton, C., 2009. A brief history of a testing time: national curriculum assessment in England 1989–2008. *Educational Research*, 51 (2), 137–159.

**Figure 1a: Kernel-density estimate of the distribution of primary school level average KS2 maths point score (2008/09 cohort, i.e. pre-boycott)**
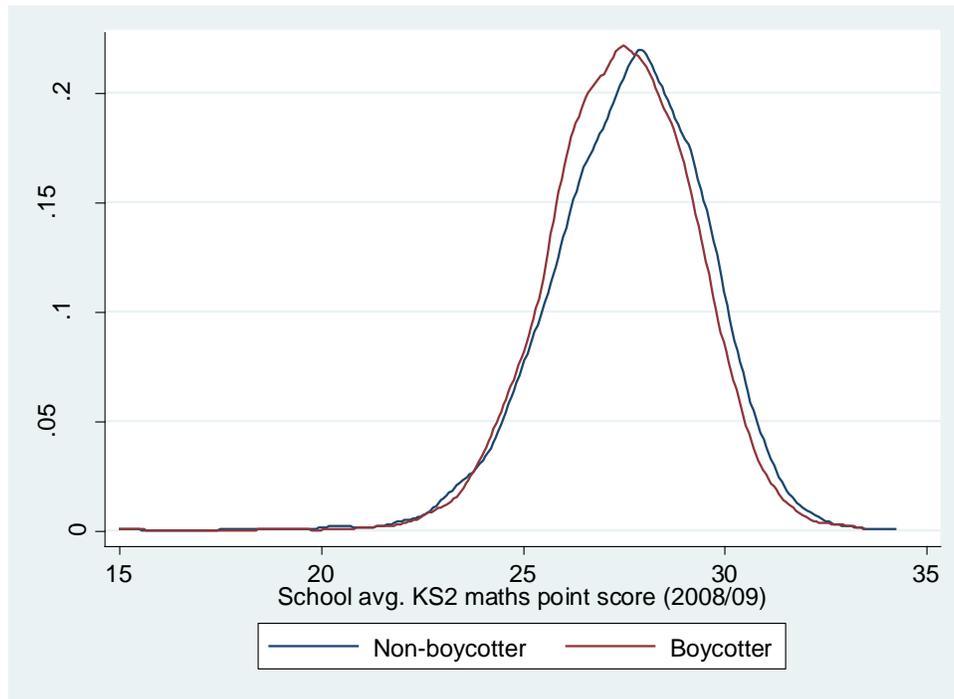


**Figure 1b: Kernel-density estimate of the distribution of primary school level average KS2 English point score (2008/09 cohort, i.e. pre-boycott)**
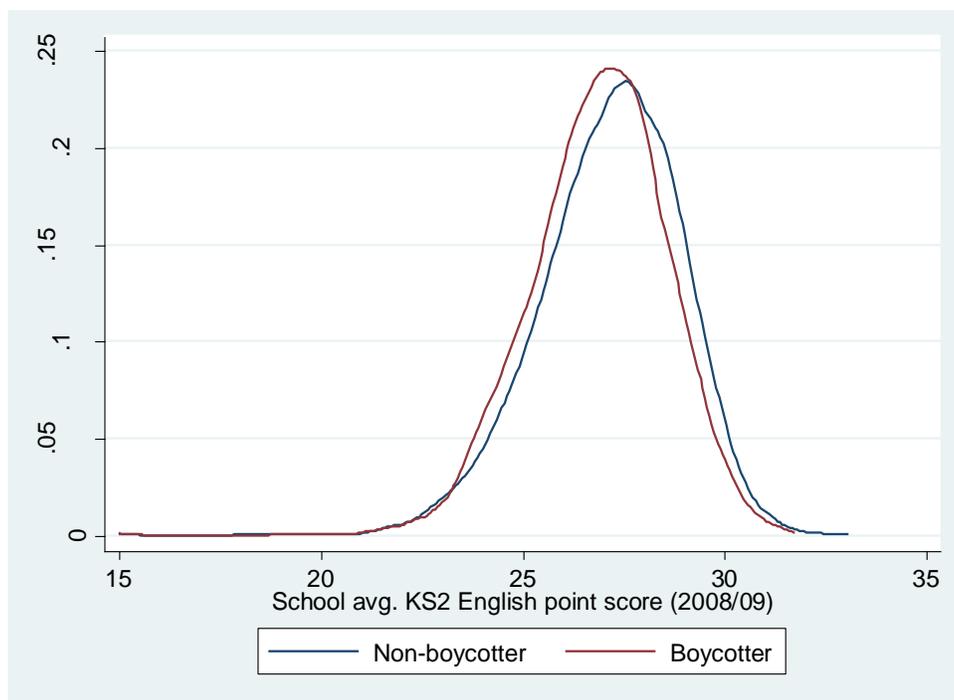
**Figure 2a: Kernel-density estimate of the distribution of primary school level average KS2 maths point score (2008/09 cohort, i.e. pre-boycott) (matched sample)**
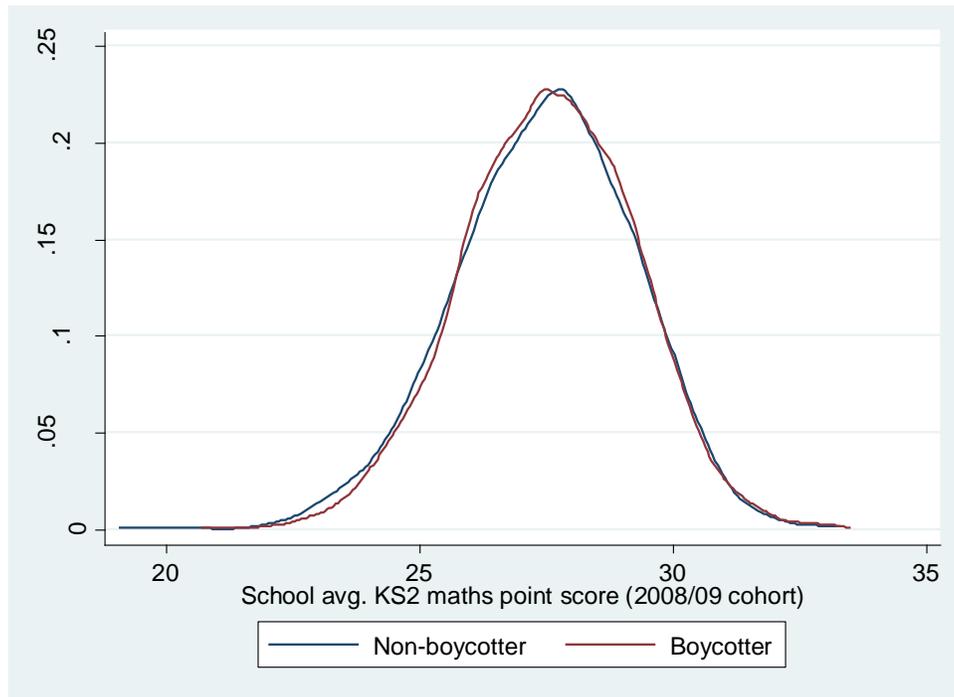


**Figure 2b: Kernel-density estimate of the distribution of primary school level average KS2 English point score (2008/09 cohort, i.e. pre-boycott) (matched sample)**
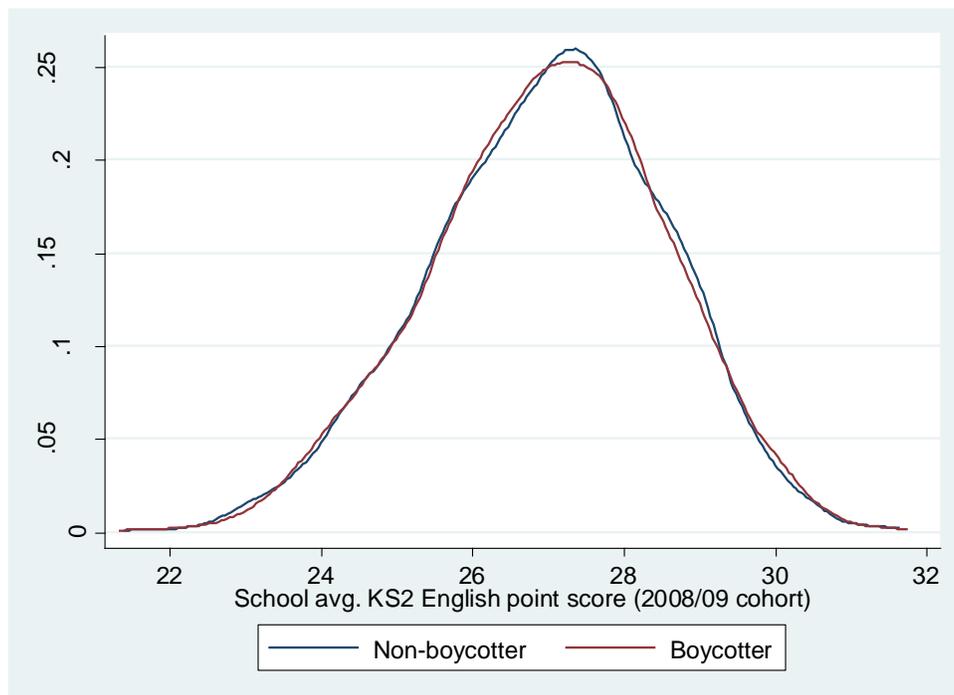
**Table 1: Expected national curriculum level attainment**

| Key stage | School year | Average age | Form of Standardised Assessment | Expected level | Highest level obtainable |
|---|---|---|---|---|---|
| KS1 | Year 2 | 7 | TAs | 2 | 3 |
| KS2 | Year 6 | 11 | Tests and TAs | 4 | 6* |
| KS3 | Year 9 | 14 | TAs | 6 | 8 |
| KS4 | Year 11 | 16 | Tests | n/a | n/a |

*Notes:* TA refers to teacher assessment. Form of standardised assessments refers to the national curriculum assessment system experienced by the cohorts studied in this paper (tests have since been reintroduced for KS1 national curriculum assessment). * between school-years 2002/03 and 2010/11 the highest obtainable level at the end of KS2 was level 5. Attainment at KS4 is assessed using nationally recognised qualifications that pupils work towards throughout KS4 (usually GCSE or vocational equivalents).

## Table 2: Primary school level means of covariates by boycott participation

|  | Non-boycotters | Boycotters | Difference (SE) |
|---|---|---|---|
| Average KS2 English point score 2008/09 cohort | 27.18 | 26.89 | 0.29*** (0.03) |
| Average KS2 maths point score 2008/09 cohort | 27.65 | 27.46 | 0.19*** (0.03) |
| Average KS1 English point score 2010 | 15.25 | 14.91 | 0.34*** (0.03) |
| Average KS1 maths point score 2010 | 15.88 | 15.55 | 0.32*** (0.03) |
| Number of pupils | 234.99 | 253.89 | -18.89*** (2.54) |
| Percentage female | 0.51 | 0.52 | -0.01*** (0.00) |
| Number of qualified teachers (FTE) | 11.47 | 12.45 | -0.98*** (0.13) |
| Pupil/teacher ratio | 20.80 | 20.99 | -0.19*** (0.07) |
| Percentage English is first language | 88.81 | 84.44 | 4.37*** (0.39) |
| Percentage white British ethnicity | 80.00 | 75.49 | 4.52*** (0.49) |
| Percentage FSM eligible | 15.55 | 19.57 | -4.01*** (0.26) |
| Percentage taking FSM | 13.56 | 17.05 | -3.50*** (0.23) |
| Percentage SEN | 20.10 | 21.08 | -0.97*** (0.17) |
| OFSTED rating, outstanding | 0.13 | 0.13 | 0.00 (0.01) |
| OFSTED rating, good | 0.50 | 0.52 | -0.02* (0.01) |
| OFSTED rating, requires improvement/satisfactory | 0.34 | 0.34 | -0.00 (0.01) |
| OFSTED rating, inadequate | 0.03 | 0.01 | 0.02*** (0.00) |
| School type, community | 0.58 | 0.67 | -0.09*** (0.01) |
| School type, foundation | 0.03 | 0.02 | 0.01** (0.00) |
| School type, voluntary-aided | 0.24 | 0.21 | 0.03*** (0.01) |
| School type, voluntary-controlled | 0.15 | 0.09 | 0.06*** (0.01) |
| Observations | 10,774 | 3,942 |  |

Notes: means are school level averages. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

**Table 3: OLS regression model for GCSE English point score**

| | (1)<br>GCSE English<br>point score | (2)<br>GCSE English<br>point score | (3)<br>GCSE English<br>point score | (4)<br>GCSE English<br>point score | (5)<br>GCSE English<br>point score |
|---|---|---|---|---|---|
| Attended boycotter school | -0.5821*** | -0.3410*** | -0.1601*** | -0.1155*** | -0.1607*** |
| | (0.0847) | (0.0472) | (0.0358) | (0.0330) | (0.0334) |
| Number of pupils | | | 0.0005*** | 0.0002* | 0.0001 |
| | | | (0.0001) | (0.0001) | (0.0001) |
| Pupil/teacher ratio | | | 0.0073 | 0.0060 | 0.0094* |
| | | | (0.0053) | (0.0050) | (0.0050) |
| Percentage English is first language | | | -0.0021 | 0.0027 | 0.0126*** |
| | | | (0.0027) | (0.0026) | (0.0027) |
| Percentage white British ethnicity | | | -0.0056** | -0.0030 | -0.0064*** |
| | | | (0.0024) | (0.0022) | (0.0024) |
| Percentage FSM eligible | | | -0.0591*** | -0.0472*** | -0.0678*** |
| | | | (0.0021) | (0.0020) | (0.0020) |
| Percentage SEN | | | -0.0171*** | -0.0075*** | -0.0172*** |
| | | | (0.0022) | (0.0020) | (0.0020) |
| Avg. KS2 English point score 08/09 | | | 0.0731*** | 0.0706*** | 0.0831*** |
| | | | (0.0156) | (0.0146) | (0.0148) |
| Avg. KS2 Maths point score 08/09 | | | 0.0431*** | 0.0359*** | 0.0577*** |
| | | | (0.0144) | (0.0136) | (0.0137) |
| Avg. KS1 point score 09/10 | | | 0.6737*** | 0.5046*** | -0.5044*** |
| | | | (0.0168) | (0.0155) | (0.0159) |
| Female | | | | 2.6593*** | 1.5492*** |
| | | | | (0.0297) | (0.0264) |
| SEN | | | | -7.4819*** | -3.4540*** |
| | | | | (0.0586) | (0.0503) |
| FSM eligible | | | | -2.7886*** | -1.6980*** |
| | | | | (0.0460) | (0.0394) |
| English is first language | | | | 0.2176*** | -1.8609*** |
| | | | | (0.0564) | (0.0542) |
| White ethnicity | | | | -0.7451*** | -0.6865*** |
| | | | | (0.0545) | (0.0464) |
| KS1 English point score | | | | | 1.2021*** |
| | | | | | (0.0060) |
| Secondary school effect | No | Yes | Yes | Yes | Yes |
| Primary OFSTED rating | No | No | Yes | Yes | Yes |
| Month of birth effect | No | No | No | Yes | Yes |
| Observations | 496,741 | 496,575 | 492,796 | 487,518 | 466,895 |
| Adj. R-Square | 0.001 | 0.160 | 0.177 | 0.305 | 0.462 |

Notes: robust standard errors clustered at the primary and secondary school level in parentheses. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

**Table 4: OLS regression model for GCSE maths point score**

| | (1) GCSE Maths point score | (2) GCSE Maths point score | (3) GCSE Maths point score | (4) GCSE Maths point score | (5) GCSE Maths point score |
|---|---|---|---|---|---|
| Attended boycotter school | -1.0404*** (0.1102) | -0.5377*** (0.0639) | -0.3039*** (0.0489) | -0.2565*** (0.0454) | -0.2910*** (0.0462) |
| Number of pupils | | | 0.0005*** (0.0002) | 0.0001 (0.0002) | -0.0001 (0.0002) |
| Pupil/teacher ratio | | | 0.0104 (0.0074) | 0.0080 (0.0069) | 0.0233*** (0.0070) |
| Percentage English is first language | | | -0.0203*** (0.0036) | -0.0042 (0.0034) | 0.0113*** (0.0036) |
| Percentage white British ethnicity | | | 0.0047 (0.0031) | 0.0060** (0.0029) | -0.0004 (0.0032) |
| Percentage FSM eligible | | | -0.0778*** (0.0028) | -0.0565*** (0.0026) | -0.0862*** (0.0027) |
| Percentage SEN | | | -0.0219*** (0.0029) | -0.0083*** (0.0027) | -0.0219*** (0.0027) |
| Avg. KS2 English point score 08/09 | | | 0.0269 (0.0206) | 0.0232 (0.0195) | 0.0792*** (0.0208) |
| Avg. KS2 Maths point score 08/09 | | | 0.1145*** (0.0195) | 0.1053*** (0.0181) | 0.1171*** (0.0186) |
| Avg. KS1 point score 09/10 | | | 0.9253*** (0.0214) | 0.7116*** (0.0199) | -0.6443*** (0.0211) |
| Female | | | | -0.9196*** (0.0344) | -0.2780*** (0.0296) |
| SEN | | | | -11.0252*** (0.0808) | -5.3805*** (0.0652) |
| FSM eligible | | | | -3.9566*** (0.0593) | -2.5293*** (0.0489) |
| English is first language | | | | -1.1946*** (0.0768) | -3.2334*** (0.0736) |
| White ethnicity | | | | -0.5707*** (0.0696) | -1.0486*** (0.0589) |
| KS1 maths point score | | | | | 1.8157*** (0.0071) |
| Secondary school effect | No | Yes | Yes | Yes | Yes |
| Primary OFSTED rating | No | No | Yes | Yes | Yes |
| Month of birth effect | No | No | No | Yes | Yes |
| Observations | 509,586 | 509,413 | 505,470 | 500,053 | 478,833 |
| Adj. R-Square | 0.002 | 0.156 | 0.172 | 0.290 | 0.508 |

Notes: robust standard errors clustered at the primary and secondary school level in parentheses. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

**Table 5: OLS regression model for 5+ GCSEs/equivalents A\*-C**

| | (1) 5+ GCSEs/ equivalents A\*-C | (2) 5+ GCSEs/ equivalents A\*-C | (3) 5+ GCSEs /equivalents A\*-C | (4) 5+ GCSEs/ equivalents A\*-C | (5) 5+ GCSEs/ equivalents A\*-C |
|---|---|---|---|---|---|
| Attended boycotter school | -0.0293*** (0.0041) | -0.0163*** (0.0024) | -0.0076*** (0.0019) | -0.0056*** (0.0017) | -0.0076*** (0.0017) |
| Number of pupils | | | 0.0000*** (0.0000) | 0.0000** (0.0000) | 0.0000 (0.0000) |
| Pupil/teacher ratio | | | 0.0008*** (0.0003) | 0.0007** (0.0003) | 0.0009*** (0.0003) |
| Percentage English is first language | | | -0.0007*** (0.0001) | -0.0001 (0.0001) | 0.0003** (0.0001) |
| Percentage white British ethnicity | | | 0.0000 (0.0001) | 0.0002 (0.0001) | 0.0000 (0.0001) |
| Percentage FSM eligible | | | -0.0032*** (0.0001) | -0.0025*** (0.0001) | -0.0035*** (0.0001) |
| Percentage SEN | | | -0.0009*** (0.0001) | -0.0004*** (0.0001) | -0.0009*** (0.0001) |
| Avg. KS2 English point score 08/09 | | | 0.0022*** (0.0008) | 0.0021*** (0.0008) | 0.0030*** (0.0008) |
| Avg. KS2 Maths point score 08/09 | | | 0.0026*** (0.0007) | 0.0022*** (0.0007) | 0.0032*** (0.0007) |
| Avg. KS1 point score 09/10 | | | 0.0322*** (0.0008) | 0.0238*** (0.0008) | -0.0239*** (0.0008) |
| Female | | | | 0.0751*** (0.0015) | 0.0392*** (0.0014) |
| SEN | | | | -0.3884*** (0.0026) | -0.1941*** (0.0026) |
| FSM eligible | | | | -0.1504*** (0.0023) | -0.0991*** (0.0020) |
| English is first language | | | | -0.0400*** (0.0028) | -0.1229*** (0.0030) |
| White ethnicity | | | | -0.0389*** (0.0026) | -0.0420*** (0.0025) |
| KS1 average point score | | | | | 0.0583*** (0.0003) |
| Secondary school effect | No | Yes | Yes | Yes | Yes |
| Primary OFSTED rating | No | No | Yes | Yes | Yes |
| Month of birth effect | No | No | No | Yes | Yes |
| Observations | 519,872 | 519,664 | 515,627 | 510,025 | 488,386 |
| Adj. R-Square | 0.001 | 0.131 | 0.147 | 0.256 | 0.376 |

Notes: robust standard errors clustered at the primary and secondary school level in parentheses. \*\*\* denotes significance at 1% level, \*\* at 5% level, \* at 10% level.

**Table 6: Heterogeneous effects models for GCSE English/maths point score**

| | (1)<br>GCSE English<br>point score | (2)<br>GCSE English<br>point score | (3)<br>GCSE English<br>point score | (4)<br>GCSE Maths<br>point score | (5)<br>GCSE Maths<br>point score | (6)<br>GCSE Maths<br>point score |
|---|---|---|---|---|---|---|
| Attended boycotter school | -0.1087*** | -0.1251*** | -0.0758 | -0.2182*** | -0.2523*** | -0.2354*** |
| | (0.0357) | (0.0354) | (0.0597) | (0.0484) | (0.0480) | (0.0787) |
| Attended boycotter school × SEN | 0.0026 | | | -0.0187 | | |
| | (0.0910) | | | (0.1172) | | |
| SEN | -3.4817*** | -3.4808*** | -3.4563*** | -5.3858*** | -5.3908*** | -5.4078*** |
| | (0.0555) | (0.0504) | (0.0515) | (0.0729) | (0.0652) | (0.0674) |
| Attended boycotter school × FSM eligible | | 0.1123 | | | 0.2067** | |
| | | (0.0796) | | | (0.1005) | |
| FSM eligible | -1.7045*** | -1.7394*** | -1.6969*** | -2.5290*** | -2.5934*** | -2.5453*** |
| | (0.0394) | (0.0466) | (0.0405) | (0.0490) | (0.0574) | (0.0507) |
| Attended boycotter school × Attended good or outstanding secondary school | | | -0.0377 | | | 0.0017 |
| | | | (0.0692) | | | (0.0914) |
| Secondary school effect | Yes | Yes | Yes | Yes | Yes | Yes |
| Primary school controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Pupil controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Pupil valued added | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 466,891 | 466,891 | 442,002 | 478,829 | 478,829 | 452,477 |
| Adj. R-Square | 0.461 | 0.461 | 0.457 | 0.507 | 0.507 | 0.503 |

Notes: robust standard errors clustered at secondary school level in parentheses. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

**Table 6 (continued): Heterogeneous effects models for GCSE English/maths point score**

| | (7) 5+ GCSEs/equivalents A*-C | (8) 5+ GCSEs/equivalents A*-C | (9) 5+ GCSEs/equivalents A*-C |
|---|---|---|---|
| Attended boycotter school | -0.0072*** | -0.0073*** | -0.0072** |
| | (0.0019) | (0.0018) | (0.0032) |
| Attended boycotter school × SEN | 0.0120*** | | |
| | (0.0038) | | |
| SEN | -0.1991*** | -0.1957*** | -0.1983*** |
| | (0.0028) | (0.0026) | (0.0027) |
| Attended boycotter school × FSM eligible | | 0.0136*** | |
| | | (0.0039) | |
| FSM eligible | -0.0995*** | -0.1037*** | -0.1008*** |
| | (0.0020) | (0.0023) | (0.0021) |
| Attended boycotter school × Attended good or outstanding secondary school | | | 0.0020 |
| | | | (0.0037) |
| Secondary school effect | Yes | Yes | Yes |
| Primary school controls | Yes | Yes | Yes |
| Pupil controls | Yes | Yes | Yes |
| Pupil valued added | Yes | Yes | Yes |
| Observations | 488,382 | 488,382 | 460,543 |
| Adj. R-Square | 0.375 | 0.375 | 0.370 |

Notes: robust standard errors clustered at secondary school level in parentheses. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

**Table 7: OLS regression model for GCSE outcomes (untreated cohort - falsification test)**

|  | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Attended boycotter school | -0.4892*** | -0.1731*** | -0.0380 | -0.0114 | -0.0405 |
|  | (0.0821) | (0.0461) | (0.0349) | (0.0323) | (0.0326) |
| Observations | 506,497 | 506,410 | 500,275 | 494,989 | 473,651 |
| Adj. R-Square | 0.001 | 0.158 | 0.182 | 0.330 | 0.482 |
|  |  |  |  |  |  |
| Attended boycotter school | -0.7776*** | -0.2589*** | -0.1025** | -0.0672 | -0.0752 |
|  | (0.1110) | (0.0639) | (0.0492) | (0.0454) | (0.0469) |
| Observations | 514,021 | 513,922 | 507,728 | 502,334 | 480,591 |
| Adj. R-Square | 0.001 | 0.153 | 0.176 | 0.311 | 0.525 |
|  |  |  |  |  |  |
| Attended boycotter school | -0.0191*** | -0.0086*** | -0.0021 | -0.0010 | -0.0022 |
|  | (0.0040) | (0.0024) | (0.0019) | (0.0018) | (0.0018) |
| Observations | 522,459 | 522,323 | 515,995 | 510,434 | 488,267 |
| Adj. R-Square | 0.000 | 0.122 | 0.142 | 0.266 | 0.377 |
|  |  |  |  |  |  |
| Secondary school effect | No | Yes | Yes | Yes | Yes |
| Primary school controls | No | No | Yes | Yes | Yes |
| Pupil controls | No | No | No | Yes | Yes |
| Pupil valued added | No | No | No | No | Yes |

Notes: robust standard errors clustered at the primary and secondary school level in parentheses. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

**Table 8: Primary school level means by boycott participation (matched sample)**

|  | Non-boycotters | Boycotters | Difference (SE) |
|---|---|---|---|
| Average KS2 English point score 2008/09 cohort | 27.00 | 27.00 | -0.00 (0.04) |
| Average KS2 maths point score 2008/09 cohort | 27.50 | 27.56 | -0.07 (0.04) |
| Average KS1 English point score 2010 | 15.04 | 15.04 | 0.00 (0.03) |
| Average KS1 maths point score 2010 | 15.68 | 15.67 | 0.01 (0.03) |
| Number of pupils | 241.54 | 243.15 | -1.61 (2.82) |
| Percentage female | 0.52 | 0.52 | -0.00** (0.00) |
| Number of qualified teachers (FTE) | 11.75 | 11.80 | -0.05 (0.13) |
| Pupil/teacher ratio | 21.00 | 21.14 | -0.14* (0.08) |
| Percentage English is first language | 87.40 | 86.23 | 1.17** (0.51) |
| Percentage white British ethnicity | 78.06 | 77.48 | 0.57 (0.65) |
| Percentage FSM eligible | 17.84 | 17.89 | -0.05 (0.32) |
| Percentage taking FSM | 15.58 | 15.57 | 0.01 (0.29) |
| Percentage SEN | 20.92 | 20.50 | 0.42* (0.22) |
| OFSTED rating, outstanding | 0.12 | 0.13 | -0.00 (0.01) |
| OFSTED rating, good | 0.48 | 0.53 | -0.05*** (0.01) |
| OFSTED rating, requires improvement/satisfactory | 0.37 | 0.33 | 0.03*** (0.01) |
| OFSTED rating, inadequate | 0.03 | 0.01 | 0.02*** (0.00) |
| School type, community | 0.61 | 0.66 | -0.05*** (0.01) |
| School type, foundation | 0.02 | 0.02 | 0.00 (0.00) |
| School type, voluntary-aided | 0.23 | 0.22 | 0.01 (0.01) |
| School type, voluntary-controlled | 0.13 | 0.10 | 0.03*** (0.01) |
| Observations | 3,486 | 3,486 | |

Notes: means are school level averages. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

**Table 9: OLS regression model for GCSE outcomes (matched sample)**

|  | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Attended boycotter school | -0.0722 | -0.1494** | -0.2193*** | -0.1522*** | -0.1965*** |
|  | (0.0970) | (0.0624) | (0.0483) | (0.0445) | (0.0453) |
| Observations | 238,425 | 238,205 | 237,881 | 235,242 | 225,402 |
| Adj. R-Square | 0.000 | 0.156 | 0.176 | 0.306 | 0.460 |
|  |  |  |  |  |  |
| Attended boycotter school | -0.3113** | -0.2901*** | -0.3909*** | -0.3163*** | -0.3528*** |
|  | (0.1273) | (0.0845) | (0.0653) | (0.0604) | (0.0603) |
| Observations | 244,975 | 244,769 | 244,434 | 241,729 | 231,580 |
| Adj. R-Square | 0.000 | 0.152 | 0.172 | 0.291 | 0.506 |
|  |  |  |  |  |  |
| Attended boycotter school | -0.0038 | -0.0091*** | -0.0133*** | -0.0102*** | -0.0127*** |
|  | (0.0048) | (0.0032) | (0.0025) | (0.0024) | (0.0024) |
| Observations | 250,133 | 249,888 | 249,545 | 246,754 | 236,412 |
| Adj. R-Square | 0.000 | 0.128 | 0.147 | 0.256 | 0.376 |
|  |  |  |  |  |  |
| Secondary school effect | No | Yes | Yes | Yes | Yes |
| Primary school controls | No | No | Yes | Yes | Yes |
| Pupil controls | No | No | No | Yes | Yes |
| Pupil valued added | No | No | No | No | Yes |

Notes: robust standard errors clustered at the primary and secondary school level in parentheses. *** denotes significance at 1% level, ** at 5% level, * at 10% level.

**Table 10: OLS regression model for GCSE outcomes (matched sample) (untreated cohort - falsification test)**

|  | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Attended boycotter school | 0.0242 | 0.0344 | -0.0626 | -0.0170 | -0.0284 |
|  | (0.0933) | (0.0611) | (0.0475) | (0.0436) | (0.0440) |
| Observations | 245,222 | 245,089 | 242,213 | 239,523 | 229,211 |
| Adj. R-Square | 0.000 | 0.153 | 0.173 | 0.323 | 0.479 |
|  |  |  |  |  |  |
| Attended boycotter school | -0.0195 | 0.0031 | -0.1501$^{**}$ | -0.0858 | -0.0846 |
|  | (0.1277) | (0.0852) | (0.0679) | (0.0625) | (0.0618) |
| Observations | 248,980 | 248,849 | 245,938 | 243,197 | 232,668 |
| Adj. R-Square | 0.000 | 0.145 | 0.165 | 0.304 | 0.523 |
|  |  |  |  |  |  |
| Attended boycotter school | 0.0073 | 0.0002 | -0.0052$^{**}$ | -0.0030 | -0.0036 |
|  | (0.0047) | (0.0032) | (0.0025) | (0.0024) | (0.0024) |
| Observations | 253,191 | 253,026 | 250,045 | 247,217 | 236,505 |
| Adj. R-Square | 0.000 | 0.118 | 0.136 | 0.261 | 0.376 |
|  |  |  |  |  |  |
| Secondary school effect | No | Yes | Yes | Yes | Yes |
| Primary school controls | No | No | Yes | Yes | Yes |
| Pupil controls | No | No | No | Yes | Yes |
| Pupil valued added | No | No | No | No | Yes |

Notes: robust standard errors clustered at the primary and secondary school level in parentheses. *** denotes significance at 1% level, ** at 5% level, * at 10% level.