Away from Home, Better at School. The Case of a British Boarding School*

Francesca Foliano[1], Francis Green[2] and Marcello Sartarelli[3]

[1]NIESR
[2]UCL Institute of Education
[3]Universidad de Alcalá

June 17, 2019

Abstract

In this paper we study whether substituting family inputs with school resources in an academically oriented environment has an impact on achievement in high-stakes national examinations. We use administrative data for England to estimate the effect of attending a selective boarding school that admits an unusually high share of pupils with low socio-economic status on attainment at the end of compulsory education. By using propensity score matching we obtain comparable control groups in selective non-boarding schools. Our main finding is that the probability of being in the top decile of achievement in the exams increases by about 18 percentage points compared to 59% for controls.

Keywords: ability, achievement gap, boarding, education, grammar school, GCSE, private school, socio-economic status, SES

JEL Classification: I21, I24, I32, J62

---

*Corresponding author: Marcello Sartarelli. Email: marcellosartarelli@gmail.com. We thank for their useful comments Pedro Albarran, Sylvain Barde, Marianna Battaglia, Arnaud Chevalier, Peter Dolton, Richard Dorsett, Iñigo Iturbe, Alfonso Miranda, Olmo Silva, Anna Vignoles, Felix Weinhardt, Yu Zhu and participants at the Universidad de Alicante Applied Seminar, University of Kent Research Seminar, AEDE 2014, SIEP 2016 and RES 2017. We also thank the management of Christ's Hospital School for kindly providing us with information about the school's admissions criteria and procedures. The responsibility for the findings in this paper are entirely our own. Financial support from the ESRC (Grant No. ESJ0191351), Spanish Ministerio de Economía y Competitividad (ECO2016-77200-P, ECO2015-65820-P and ECO2013-43119) and Universidad de Alicante (GRE 13-04) is gratefully acknowledged.

1 Introduction

Gaps in achievement by socio-economic status (SES) are a global policy challenge (see for a review Sirin, 2005; Reardon, 2011). In England these gaps are observed as early as the primary education phase and usually do not change over time or tend to increase as students progress through the education system (Dearden *et al.*, 2011). While gaps by SES tend to be concentrated among pupils who are initially low achievers, they have also been found among high achievers (Crawford *et al.*, 2014; Jerrim, 2017). This may have high opportunity costs if pupils who have the potential to perform well at school are held back or slowed down by the environment in which they have grown up.

Most policies designed to counteract the influence of deprivation on pupil achievement are based on the assumption that increasing school inputs boosts academic outcomes. However, the effect of these policies may be confounded by parental responses to the change in school inputs (Todd and Wolpin, 2003, 2007): family inputs may reinforce the role of school policies if parents respond by investing more into their children's development or dilute it if they invest less.[1]

Boarding schools offer the opportunity to observe a context where family inputs are substituted for school inputs, i.e. they reduce the role of family inputs for all pupils, since they offer education during the day and lodging at night. However, obtaining clean estimates of the effect of attending a boarding school on pupil achievement is an empirical challenge as it may be confounded by a selection effect if boarding school pupils and pupils in other schools differ substantially in ability or family resources. This problem has been circumvented in recent studies which exploit lotteries to oversubscribed boarding schools to estimate the clean effect of these types of educational institutions on achievement. Randomly admitted pupils obtain substantially higher test scores than non-admitted ones in boarding schools in poor neighbourhoods in the US (Curto and Fryer Jr, 2014). Related research exploiting random admission in an elite school in France obtains similar results (Behaghel *et al.*, 2017).

This paper is the first to study the effect of boarding education in England. The aim of our empirical analysis is to investigate the effect of boarding education

[1] Recent examples of school policies include the introduction of sponsored Academy schools in disadvantaged areas in the UK (Eyles *et al.*, 2016) and of Charter schools in the US (see a review in Epple *et al.*, 2016), as well as more narrowly targeted interventions in urban schools in the UK, such as Excellence in the Cities (Machin *et al.*, 2004).

by studying a selective boarding school called Christ's Hospital (CH hereafter). This school admits talented pupils from different backgrounds and funds the cost of their education with a variety of means-tested bursaries. For our analysis we use rich administrative data of pupils in England and measures of achievement at the ages of 7, 11 (Key Stage 1 and 2) and the results of compulsory school final exams at age 16 (General Certificate of Secondary Education, GCSE hereafter) for five consecutive cohorts of pupils. We use propensity score matching to find, for each pupil at CH, a pupil in a selective day school who is as similar as possible in observable characteristics.

CH is financially supported by a wealthy foundation which funds means-tested bursaries for the majority of its pupils and is devoted to helping high achievers with low SES. We assess the effect of boarding by comparing CH pupils to those in other selective schools that do not provide boarding (i.e. selective day schools). Our first control group includes pupils who attended grammar secondary schools and whose primary school was in the same local authority as those attended by pupils who then went to CH. This ensures that both school and non-school environment are comparable for CH and control group pupils. While grammar schools are academically selective like CH, they differ in that they do not offer boarding and have substantially fewer resources. For the second control group we select pupils from independent schools, which are often as well-resourced as CH but tend to be less academically selective. CH pupils have lower SES and somewhat higher achievement in primary school tests than pupils in the two control groups. The characteristics of its pupils make CH an outlier in English private education. This confirms the importance of using rich measures of primary school achievement and SES in our empirical design to estimate a clean effect of attending CH.

We find that the achievement of pupils attending CH is significantly higher at GCSE than for pupils in either grammar or independent day schools. The probability of at least five GCSEs at A-A* (i.e. of being in the top two deciles in the distribution of the number of GCSEs at A-A*) is 18 percentage points higher or 30% relative to the mean value for the control groups. We assess whether there is heterogeneity in our main results and we find that the effect for girls is slightly higher relative to boys although the difference is not significant. Similarly, we detect no significant difference between pupils with high and low SES, which suggests that the effect is not concentrated on high-SES students, but is equally large among low-SES students. Crucially, predetermined characteristics for pupils at CH and for controls groups in grammar and independent schools are balanced. Our main results are robust to a placebo test assessing whether an extensive set of predetermined outcomes is systematically different for pupils at CH and pupils in grammar or independent schools. They are also robust to a sensitivity analysis assessing the bias of the main results in the presence of selection on unobservables.

Our paper offers a proof of concept that a boarding secondary school admitting high ability pupils with a lower SES than pupils in other selective day schools can improve their achievement. Our results contribute to the school choice literature by suggesting that currently available alternatives to standard schooling options may play a role in reversing the achievement gap at the end of compulsory schooling for disadvantaged pupils. Our results also contribute to the recent policy debate over the use of boarding schools for disadvantaged children in England (Department for Education, 2014, 2016), a debate which So far has not been informed by a thorough quantitative.

The rest of the paper is structured as follows. Section 2 reviews the relevant literature. Section 3 describes the institutional setting for compulsory education in England and the data that we use in the empirical analysis. Section 4 outlines the econometric strategy. Section 5 describes the main results, section 6 reports the results of a sensitivity analysis and section 7 concludes.

2 Literature review

In this section we describe the findings of the literature in economics on the effects of boarding on academic and non-academic outcomes. Boarding secondary schools for bright pupils with low SES have been introduced in recent years in the US. SEED schools in Washington and Baltimore are the only urban public schools that combine the charter school model with a 5-day-a-week boarding program in poor neighbourhoods. Curto and Fryer Jr (2014) estimate the impact of attending SEED schools, i.e. the joint effect of a charter school and a boarding school, on achievement by exploiting a lottery-driven admission system that is used when a school is oversubscribed. They compare the achievement of students admitted and rejected by the lottery and find that achievement in SEED schools is higher by about 20% of a sd in reading and maths, with the results being mainly driven by females

In France public 'boarding schools of excellence' for poor and high achieving pupils have been opened in deprived suburbs of large French cities. Behaghel *et al.*

(2017) exploit an admission lottery to study the effect of attending one such school in the suburbs of Paris. They find that by the end of the first year, achievement in French and in maths is lower (by 6.5% and 3.7% of a standard deviation, s.d. hereafter) although these differences are not significant. A subjective measure of well-being, obtained by way of a survey, is also weakly significantly lower (29.8% s.d.) and is driven by frictions in adapting to the boarding environment. In contrast, after the second year maths scores are significantly higher (28% s.d.) while they are lower in French (11.5% s.d.) although this difference is not significant. Well-being is also significantly higher than in the first year (11.8% s.d.), driven by significantly higher scores to the question on whether children feel at home. Improvement in maths is driven by those students who were in the top three deciles of the distribution of maths scores when they enrolled.

Curto and Fryer Jr (2014) and Behaghel *et al.* (2017) quantify the effect of boarding school on achievement for low SES pupils by exploiting admission lotteries. The authors obtain clean estimates of the effect of boarding by using this quasi-experimental setting. However using lotteries presents a drawback: since oversubscribed schools are in higher demand than others that are not oversubscribed, their quality is likely to be higher because they may have more resources or more motivated and better qualified teachers. As quality is unobserved, estimates of the boarding school effect obtained by exploiting lotteries may be upward biased. However, note that this limitation applies also to our setting because CH is oversubscribed.

In a study based on observational data Andersson and Johansson (2013) investigate the effect of boarding primary education in a rural Swedish county in the 1940s. Pupils living far away from primary schools were allowed to board accommodation where they were given food and lodging and out-of-school time was dedicated to work, tidiness and obedience. These pupils were more likely to come from a less advantageous background than the rest of pupils enrolled in local primary schools. Andersson and Johansson (2013) find that boarding has a positive effect on achievement at the end of primary school and this effect increases monotonously with the number of semesters pupils had boarded. However, they find no effects of boarding on a set of mid to long-term outcomes such as years of education and earnings. Andersson and Johansson (2013); Curto and Fryer Jr (2014); Behaghel *et al.* (2017) are the only studies in economics, to the best of our knowledge, that investigate the effects of boarding education.[2]

Overall, the evidence summarised in this section shows that disadvantaged pupils in boarding schools tend to achieve more and that this seems to be driven by higher motivation and study effort. Our study contributes to the literature by adding evidence for England and showing that boarding education has a positive effect on the achievement of pupils from different backgrounds.

3 Institutions and data

We describe the English education system in section 3.1, our treatment and control groups in section 3.2 and 3.3 and our data in section 3.4.

3.1 National curriculum

There are 11 years of compulsory education in the English state school system, divided into primary and secondary phases and four Key Stages. Primary school starts with Key Stage 1 (age 5 to 7) and is followed by Key Stage 2 (age 7-11). Secondary school starts with Key Stage 3 (age 11 to 14) followed by Key Stage 4 (age 15-16). All Key Stages end with a national standardised assessment that is based either on an evaluation carried out by teachers, at the end of Key Stage 1 and 3, or on the results of externally marked tests, at the end of Key Stage 2 and 4. We present this system in Table 1.[3]

The National Curriculum in England sets targets that pupils are expected to achieve in each subject and each Key Stage. These targets are expressed as levels that range between 1 and 8 on an integer scale and define a precise set of skills acquired by a pupil by the end of the Key Stage; a list of the expected level for each Key Stage is reported in the final column of Table 1. An interesting feature of this system is that at the end of Key Stage 2 pupils are assessed twice: first by their own teachers based on their daily interaction over the academic year; and second by external examiners who mark their exam papers.[4]

[2]Boarding schools have been studied in psychological research with a focus on their consequences on pupil well-being, with mixed findings (Lester *et al.*, 2015; Wires *et al.*, 1994; Fisher *et al.*, 1986; Martin *et al.*, 2014; Hodges *et al.*, 2016; Schaverien, 2004, 2011).

[3]GCSE exams are taken in the final year of compulsory education, at the end of Key Stage 4. A single regulator ensures that the same standards are applied across different exam boards and over time. (Machin *et al.*, 2018).

[4]Key Stage 2 tests are marked using an integer score from 0 to 100. Targets are cutoff values in test scores that are set out to help pupils, parents and schools interpret progress throughout

Table 1: Compulsory education in England

Phase	Age	School year	Key Stage	Assessment	Expected achievement level
Primary School	5-7	1-2	1	Teachers (state schools)	2
	7-11	3-6	2	External & Teachers (state schools)	4
Secondary School	11-14	7-9	3	Teachers (state schools)	5 or 6
	15-16	10-11	4	External (GCSE) (all schools)	5 GCSEs at A*-C

3.2 Christ Hospital

CH is an independent selective and boarding-only mixed school located in West Sussex, South-East England. It is a Christian institution which according to its mission statement is dedicated to providing a stable background and boarding education of high standard to 830 boys and girls, having regard especially to children of those families in social, financial or other particular need. Thanks to its own financial endowment, it can pay over 80% of the costs of its pupils' education. In addition, these resources permit a large selection of optional subjects to be offered in the academic curriculum, and pupils become involved in a broad range of extracurricular activities including music, art, drama, public speaking, community action and sport, making use of the extended non-teaching time available in a boarding environment.[5]

Applicants to CH have to meet its academic standards and also be judged suitable to board. They are expected to be working towards level 5 at Key Stage 2 in English, Maths and Science. After a first selection based on school reports, successful applicants are invited in for an initial assessment in English and Maths. Those who pass are invited to a second assessment stage consisting of additional English and Maths tests a few months later and also to stay in the school overnight to help the school assess their suitability to board. Calculations from CH show that each assessment stage screens approximately 50% of all applicants. Overall, achievement at Key Stage 2, SES and suitability to board are CH admission criteria.[6]

compulsory education.

[5]The name of Christ Hospital School is also used in the empirical analysis in compliance with guidelines on disclosure control that can be found in point 9.5 in the National Pupil Database Agreement for the supply of data and after obtaining written approval from the Department for Education.

[6]Anecdotal evidence suggests that CH relies mainly on word of mouth by its alumni for pub-

3.3 Control groups

We study the effect of boarding education on achievement by comparing pupils attending CH, our treatment group, to pupils in two control groups selected from grammar schools and non-boarding independent schools. Mixed grammar schools are highly selective, academically oriented for historical reasons and include different school types. In our data about 54% are Foundation and 24% are Voluntary Aided or Voluntary Controlled, which are not controlled by Local Authorities (LAs). The remaining 22% are Community grammars, which are not independent of control from LAs.[7] Independent schools are fee-paying private institutions attended by about 7% of all pupils, many of these are boarding schools. They set their own examinations at the end of each academic year and the only national assessment their pupils sit during compulsory schooling is GCSE. Independent schools admit small number of pupils on means tested bursaries: we expect to find most of our matched controls from the small number of pupils that independent schools admit on means-tested bursaries.[8]

We define our first control group of pupils from selective day schools using two steps. First we include all pupils who attended grammar schools. We choose these schools because they are similar to CH in that they are academically selective, but differ because they have substantially fewer resources. Then we select those grammar school pupils who went to primary schools in the same local authorities as those attended by pupils who went on to attend CH. We do this to ensure that the CH pupils and those in the control groups face the same choice set of secondary schools, live in the same geographical area and have experienced the same local government.

We start building our second control group using all pupils in independent

licity.

[7]There are six types of state schools in England which differ for their degree of autonomy from the LA and by type of governance: community schools, voluntary controlled schools, foundation schools, voluntary aided schools, city technology colleges and academy schools. Between the non-academies the two extremes in terms of autonomy are Community schools, run predominantly by the LA, and Voluntary Aided schools managed by their governing body. Academies are completely independent from LAs and receive their fundings directly from the central government.

[8]The percentage of pupils attending independent schools varies between about 5% for pupils aged 5-10, 8% for those aged 11 to 15, and 18% for those aged 16 to 18. About 13.5% of pupils are boarders in independent schools and only 1% of all independent schools has only boarding pupils. The average termly boarding fee is 8,780 pounds while the average termly day fee is 3,903 pounds. Bursaries, scholarships and discounts are available: around 8% of pupils have received means-tested bursaries and 1% of all pupils paid no fees at all (Independent Schools Council, 2014).

schools, which are academically selective to varying degrees and have far more resources than state schools (Green *et al.*, 2012). From that pool we select those pupils who attended schools which have a Christian foundation, like CH, and who were in primary school in the same LAs at those attended by CH pupils. Pupils from grammar or independent Christian schools who were in a primary school in the same LAs as those attended by pupils at CH are approximately 10% of all pupils in grammar and in independent schools.

Figure 1: Number of CH pupils by local authority where they went to primary school

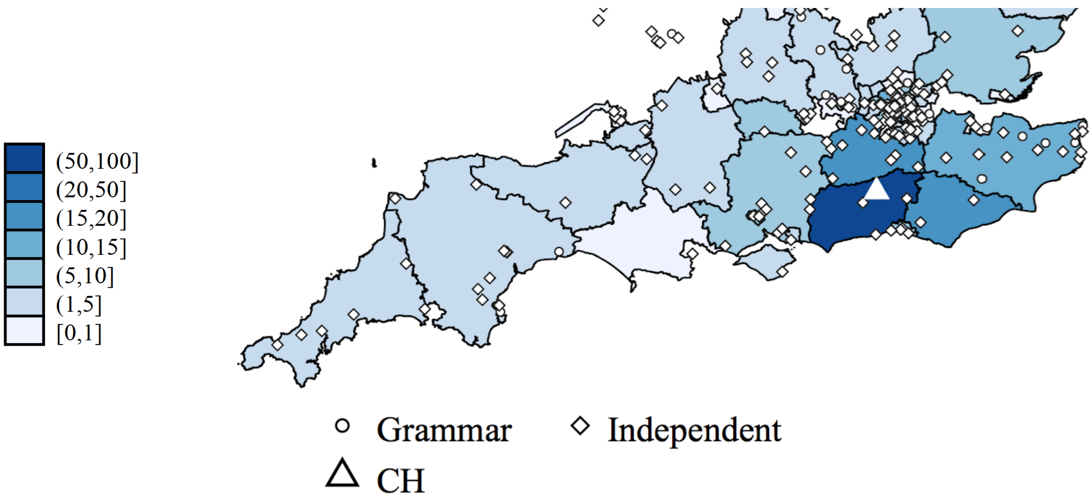


Figure 1 shows the numbers of CH pupils who went to primary school in each LA in southern England. The location of CH is marked using a white triangle. It also shows the set of all grammar and independent secondary schools attended by pupils who were in a primary school located in the same LA as those attended by CH pupils, marked using circles and diamonds respectively. LAs from which no pupil goes to CH after completing primary school are not shown. The map shows that grammar schools are fewer relative to independent schools and the closest grammar is located further away from CH than the closest independent school.[9]

Grammar schools are a type of state school, so they are funded by government through the LAs. Independent schools receive no direct government funding, though about 80% of them are constituted as charities receiving significant tax exemptions (Independent Schools Council, 2014). They receive most of their income in the form of fees. Table 2 shows proxies of teaching resources separately for CH, for

[9]The full map can be found in Figure A.1 in the Appendix. It reports grammar and independent schools separately by whether they were attended by pupils similar to those at CH, i.e. matched controls, based on the propensity score that will be defined in section 4.

Table 2: Resources in different types of schools

	CH	Grammar	Independent	State
Pupil/teacher ratio	8.80	16.44	7.91	14.65
Pupil/Full-time qualified teachers ratio	9.91	18.87	6.80	13.89
Pupil/Part-time qualified teachers ratio	50	62.50	8.40	38.46

our control groups and for state schools derived from the school-level data provided by the Department for Education (School Workforce Census for the school year 2006/07). The teaching resources of CH are similar to those of independent schools. However, CH has a substantially lower pupil/teacher ratio, defined both using the total number of teachers and the number of full-time qualified teachers, relative to grammar schools. The table shows that both CH and independent schools have higher resources relative to state schools and that grammar schools have similar resources to state schools. Finally the table shows that CH and grammars do not use part-time teachers very much compared to independent schools.

3.4 Data

Our analysis is based on individual-level administrative data of pupils in England called the National Pupil Database (NPD), which integrates information on a wide range of socio-demographic characteristics with detailed assessment records. The final dataset of about 2.5 million pupils contains information on five cohorts who attended primary state schools and sat their Key Stage 2 tests in years 2002-2006 and GCSEs at the end of Key Stage 4 in years 2007-2011. Out of all pupils in the data, 429 went to CH after completing primary education in state schools, an average of 86 pupils each year. About 70,000 went to secondary grammar schools and about 80,000 to independent ones.

We match pupils at CH with pupils in grammar or independent schools using achievement at Key Stage 1 and 2 and additional predetermined characteristics. We include Key Stage 2 test scores to ensure that CH pupils and matched controls have similar achievement at age 11. We do this because selective secondary schools screen pupils almost exclusively based on their attainment at this stage of education.[10] In

[10]Pupils who want to attend selective schools in England are usually required to sit specific examinations, such as the 11 Plus, during the last year of primary school. The results of these examinations are not publicly available and they cannot be used in our empirical analysis. However the correlation between these additional test results and Key Stage 2 exam scores is likely to be high as both examinations test curriculum skills.Allen and Bartley (2017) finds that the English

addition Key Stage 2 tests are externally marked and therefore less likely to be affected by teachers' biases. We also include teacher assessments at Key Stage 1 to ensure that CH pupils and matched controls have similar initial achievement at age 6-7.

We use the following socio-demographics in matching pupils from CH with those in grammar and independent schools: gender, ethnicity and two proxies for SES. The first of these proxies is the income deprivation affecting children index (IDACI), measuring the share of children in low income households by local area. The second is a dummy equal to 1 if a pupil is eligible for Free School Meals (FSM) because her parents receive some form of income support. We also consider distance from CH, defined as the distance between the primary schools attended by pupils in year 6 and CH.[11]

Finally we create proxies for motivation, a specific type of non-cognitive skill, by exploiting information on pupil achievement by subject at Key Stage 2 from two comparable sources, teacher assessment (TA) and test scores. TAs are carried out before teachers know their pupils achievement in tests. We start by creating a dummy equal to 1 if TA by Key Stage 2 subject is greater than 4, the achievement level at Key Stage 2 which the Department for Education expects from pupils at the end of Key Stage 2. Teachers observe their pupils on a daily basis over the academic year therefore their assessment includes both a measure of cognitive skills, such as logic and numeracy, and a consideration of their non-cognitive skills, such as motivation. These two measures are typically hard to disentangle for teachers. National test scores are instead predominantly a measure of cognitive and curriculum skills. A TA greater than the level achieved in a given national test is an indicator of high motivation as the teacher has formed a belief on that pupil overall attainment, that includes both curriculum skills and attitude towards learning, and this is higher than the clean measure of achievement as Key Stage 2 test. Therefore we define our measure of motivation as a dummy equal to 1 if the difference between a pupil's TA and the achievement in the national test is positive in at least one subject. We report this measure of motivation in the summary statistics in Table 3 and we use

element of the 11 Plus shows a correlation of 0.62 with reading score at Key Stage 2; similarly, the maths element is correlated at 0.68 with maths at Key Stage 2.

[11] Our choice to use FSM and IDACI as joint proxies for low SES is supported by results in Crawford and Greaves (2013) showing that a novel dichotomous measure of educational deprivation based on rich survey data correctly classified as deprived 80% of children with FSM status and 72% of children in the top quartile of the IDACI distribution.

it in the estimation of the pscore.

TAs are richer measures of pupil achievement than test scores because teachers observe their pupils on a daily basis. Although teachers have to support their TAs with examples and by answering detailed questions provided by the Department for Education to ensure that their assessments are in line with national standards, these assessments are more subjective than externally graded tests. We argue that the dummies we created to capture whether TAs and test levels differ, measure specifically non-cognitive skills such as motivation, whereas teacher bias is a relatively minor issue in a setting of high ability pupils.[12]

Table 3: Summary statistics

	CH	Grammar-CH		Independent-CH		All schools-CH	
		Diff.	p-value	Diff.	p-value	Diff.	p-value
Female	0.443	0.033	0.213	-0.080	0.002	0.048	0.044
White	0.599	0.210	0.000	0.187	0.000	0.221	0.000
FSM	0.146	-0.111	0.000	-0.112	0.000	-0.001	0.939
IDACI	0.227	-0.079	0.000	-0.101	0.000	-0.016	0.101
IDACI > median	0.629	-0.137	0.000	-0.217	0.000	-0.001	0.974
KS1 Eng reading DfE	0.126	-0.134	0.006	-0.499	0.000	-0.932	0.000
KS1 Eng writing DfE	0.184	-0.195	0.000	-0.516	0.000	-0.850	0.000
KS1 Mat DfE	0.117	-0.127	0.017	-0.454	0.000	-0.935	0.000
KS2 Eng score	75.801	-2.702	0.000	-8.566	0.000	-18.986	0.000
KS2 Maths score	88.361	-1.366	0.002	-10.051	0.000	-25.002	0.000
KS2 Sci score	70.653	-1.631	0.000	-5.179	0.000	-13.024	0.000
KS2 Eng lev. >4	0.857	-0.106	0.000	-0.288	0.000	-0.558	0.000
KS2 Mat lev. >4	0.894	-0.033	0.042	-0.263	0.000	-0.582	0.000
KS2 Sci lev. >4	0.928	-0.055	0.000	-0.197	0.000	-0.481	0.000
1+ KS2 TA gt tests levels	0.162	0.044	0.026	0.035	0.070	0.060	0.001
Miles to CH	31.749	30.575	0.000	8.371	0.000	0.000	.
Miles to closest grammar	1.932	-0.891	0.000	0.377	0.040	0.000	.
N	377	6,421		7,183		2,713,111	.

The first column of Table 3 shows summary statistics for the most important observable characteristics that we use to match CH pupils with pupils in our control groups. The central columns show differences between our control groups and CH, along with p-values of the null hypothesis of no difference. Finally, the two last columns on the right-hand side show differences between the population of pupils in

[12] In contrast (Burgess and Greaves, 2013) look at the whole population of pupils, with greater heterogeneity by socio-demographics and achievement, and show that teachers may exhibit a bias in their TAs in favour of certain groups of students.

all schools in England and CH. The table shows that the share of females is similar across groups, though somewhat lower for independent schools, while the share of white pupils is lower at CH. We find that IDACI and FSM is higher at CH relative to grammar and independent schools when we look at proxies for SES.

Table 3 shows that achievement is higher for CH pupils relative to all other groups. This holds both for achievement in Key Stage 1 tests, which we report rescaled following the conversion table in Department for Education (2017), and in Key Stage 2 tests. The difference in attainment across schools appears more noticeable when we compare shares of students scoring in tests above level 4, the expected level at Key Stage 2. The share of pupils whose TA in at least one Key Stage 2 subject is greater than the test level achieved in the same subject varies from 16% to 23% for our full data sample. Finally, for pupils attending grammar schools the distance to CH is twice that for CH pupils while the distance to CH is only slightly higher for those pupils attending independent schools, as shown in Figure 1. Overall, Table 3 suggests that to obtain clean estimates of the effect of on achievement at GCSE, we need to compare CH pupils with pupils in grammar and independent schools with similar SES and achievement in Key Stage 1 and 2 tests.[13]

We create the following three outcomes of interest for our empirical analysis: a dummy equal to 1 if the number of GCSEs at A is at least 1; a dummy equal to 1 if the the number of GCSEs at A* is at least 1; a dummy equal to 1 if the number of GCSEs at A or A* is at least 5. These outcomes are typically good predictors of the decision to enrol in post-compulsory education (Chowdry *et al.*, 2013). Figure A.2 in the Appendix, which reports histograms of the continuous variables underlying our outcomes of interest at GCSE separately for CH, grammar and independent schools and for our full dataset with about 2 million pupils, shows that CH pupils do better on average and are over-represented among the top achievers with 5 or

[13] Distances are computed by using publicly available data on school postcodes and on longitude and latitude coordinates associated to postcodes, measured by using the World Geodetic System 1984 (Ordnance Survey website). These information is then converted into Ordnance Survey Maps northing and easting coordinates thanks to a Helmert transformation (Watson, 2006) to eventually obtain distances in miles. We use the postcode of each pupils' primary schools rather than that of their home as the latter information is sensitive. We argue that our results would not change substantially after obtaining pupils' postcodes to compute a more precise measure of the distance between home and secondary school, as anecdotal evidence suggests that distance to primary school tends to be typically low and it seems to be subject to moderate variation across pupils (Burgess *et al.*, 2015).

more GCSEs at A or A*.[14]

4 Econometric strategy

We estimate the effect of going to CH on achievement in the compulsory school leaving exams by using propensity score (pscore) matching, an econometric strategy based on selection on observables. This is possible thanks to the unique admission criteria based jointly on merit and on SES and to the rich set of pupil observable characteristics in the administrative data.

$$\Delta^{ATT} \quad = \quad E[A(1) - A(0) \mid D = 1] \tag{1}$$

Let D be a dummy indicating whether pupils go to CH, with $D = 1$ for pupils at CH (treatment) and $D = 0$ for those in a selective day school (controls). Let also $A(1)$ and $A(0)$ be the potential outcomes, i.e. achievement, for treated and for controls. Let X be a set of predetermined observable characteristics for pupils. Our parameter of interest is the average treatment on the treated (ATT), which we denote Δ^{ATT} and define in our setting as the mean effect of attending CH, i.e. the treatment group, rather than a selective day school, i.e. the control group, as shown in equation (1).

To recover via the law of iterated expectations the unobservable term $E[E[A(0) \mid D = 0] \mid D = 1]$ in equation (1) we rely on the assumption that admission to CH depends only on observables, also known as selection on observables or Conditional Independence Assumption (CIA). Under this assumption assignment to the treatment or to the control group is independent on the treatment D conditional on the set of observables X, formally $A(1), A(0) \perp D \mid X$. However, when the number of observable characteristics in the vector X is high, it may not be possible to find for some pupils at CH pupils in control schools with the same observables X, unless the number of observations in the data is very high. This problem, known as curse of

[14] In choosing our outcomes of interest we focused on the highest grades in GCSEs, i.e. A or A*, since all secondary schools we consider are selective and those pupils tend to achieve towards the high end of the distribution of grades in GCSEs. We did not choose the probability of achieving five or more GCSEs at A*-C, a lower grade as that is about 98% in selective schools and, similarly, the mean number of GCSEs taken by pupils in these schools is 10 and shows little variation across schools. Achievement in English and Maths at GCSE are not used as outcomes as this information is not available for CH and for some independent schools in NPD data.

dimensionality, is solved by using the probability of going to CH given observable characteristics X or pscore, i.e. $P(D = 1 \mid X)$.

We ensure that for each pupil at CH there is one or more with very similar observables in the control group by imposing the common support (CS) condition, i.e. $0 < P(D = 1 \mid X) < 1$. Finally, after estimating the pscore with a logit model, we match treated pupils with very similar pupils from the control group by using the nearest neighbour matching method. We obtain two sets of estimates by using two different control groups: grammar and independent schools. While in our preferred specification we use the nearest neighbour method to match pupils, we also assess the sensitivity of our results to using different matching methods based on the pscore.[15]

The assumption we have made in our analysis so far was that the choice faced by talented pupils was binary: either CH or another type of selective school, such as an independent school. However, at the end of primary school a talented pupil may have been granted admission to CH as well as to a grammar and an independent secondary school. This characteristic can be accounted for by extending the binary propensity score matching framework to the case of multiple treatments thanks to the matching estimator proposed in Lechner (2002). By allowing multiple treatments, the treatment variable D in our setup is no longer binary and can take multiple values instead. In our setup of secondary school choice, D is equal to 0 if a pupil chooses an independent school (which we set as the baseline) although this choice does not affect results, to 1 if the choice is a grammar one and to 2 for CH.

We first estimate a multinomial logit model of school choice using the set of observables X as covariates. We then compute the predicted probabilities $\hat{P}^j(X) = \hat{P}(D = j \mid X)$ of attending an independent school ($j = 0$), a grammar school ($j = 1$) or CH ($j = 2$). To estimate the effect of attending CH relative to, for example, an independent school we compute the conditional probability $\hat{P}^{2|2,0}(X) = \frac{\hat{P}^2(X)}{\hat{P}^2(X)+\hat{P}^0(X)}$. Finally, we use the estimated conditional probability in Lechner (2002) as a balancing score in a matching estimator setting with multiple treatments to estimate the unobserved term $E[E[A(0) \mid D = 0, P^{2|2,0}] \mid D = 2]$, i.e. to match pupils at CH ($D = 2$) and pupils in independent schools ($D = 0$) with very similar values of the conditional probability $P^{2|2,0}$. We repeat the procedure to estimate the effect of attending CH relative to a grammar school.[16]

[15] ATT estimation with binary treatment was conducted using the Becker and Ichino (2002); Leuven and Sianesi (2015) software routines.

[16] ATT estimation with multiple treatments was conducted by implementing the algorithms

5 Results

We begin this section by showing the estimates of the pscore and means of predetermined characteristics separately for pupils at CH and for those in the control groups in subsection 5.1. We then show propensity score matching estimates of the effect of going to CH on achievement in the compulsory school final exam in subsection 5.2. We present a sensitivity analysis in section 6.

5.1 Propensity score and balance of predetermined characteristics

We estimate the propensity score by using a logit model and the following predetermined characteristics as independent variables: achievement levels in Key Stage 1 English and Maths tests rescaled following Department for Education (2017); scores in Key Stage 2 tests; dummies for whether achievement levels by subject at Key Stage 2 are greater than 4; gender and white ethnicity dummies; a dummy for FSM; a dummy equal to 1 if IDACI is above the median; measures of distance. We also create proxies for non-cognitive skills, such as motivation and resilience, by exploiting information on pupil achievement by subject at Key Stage 2, as described in section 3.4. The measure of non-cognitive skills we use in the pscore estimation is a dummy equal to 1 if Key Stage 2 teacher assessment is greater than test achievement in at least one of the three subject tests.

We select the richest possible set of predetermined characteristics that were balanced in each pscore block determined by the matching algorithm we use (Becker and Ichino, 2002). The advantage of estimating the pscore by using five cohorts of data is that larger samples improve the quality of the matching between CH pupils and pupils with very similar observable characteristics in the control groups. The full list of variables can be found in Table 4, along with descriptive statistics. The table is divided in two vertical panels, with the panel on the left and right showing means for predetermined characteristics used when the control group is pupils in grammar and independent schools respectively. For each predetermined characteristic we report the mean for those pupils at CH for whom the matching algorithm has found a match, i.e. match treated pupils, for matched controls and for non-matched controls.

Table 4 shows some differences between control groups in the predetermined

proposed in Gerfin and Lechner (2002); Lechner (2002); Frölich *et al.* (2004).

Table 4: Mean of controls for CH pupils and (non-)matched controls

	Grammar				Independent			
	Used in pscore	Matched CH	Matched controls	Non-matched controls	Used in pscore	Matched CH	Matched controls	Non-matched controls
Female	Yes	0.443	0.430	0.478	Yes	0.443	0.437	0.359
White	Yes	0.599	0.632	0.819	No	0.599	0.681	0.791
FSM	Yes	0.146	0.148	0.029	Yes	0.146	0.124	0.030
IDACI	No	0.227	0.186	0.146	Yes	0.227	0.204	0.122
IDACI > median	Yes	0.629	0.585	0.486	No	0.684	0.637	0.459
KS1 Eng reading DfE	Yes	0.126	0.089	-0.014	No	0.525	0.495	0.003
KS1 Eng writing DfE	Yes	0.184	0.175	-0.021	Yes	0.533	0.525	-0.009
KS1 Mat DfE	No	0.117	0.111	-0.017	Yes	0.477	0.480	-0.001
KS2 Eng score	Yes	75.801	75.113	72.988	Yes	75.801	76.625	66.770
KS2 Maths score	Yes	88.361	88.264	86.925	Yes	88.361	87.903	77.835
KS2 Sci score	Yes	70.653	70.466	68.941	Yes	70.653	70.546	65.222
K2 Eng lev. >4	Yes	0.857	0.810	0.748	Yes	0.857	0.841	0.555
K2 Mat lev. >4	Yes	0.894	0.902	0.858	Yes	0.894	0.870	0.619
K2 Sci lev. >4	Yes	0.928	0.917	0.871	Yes	0.928	0.900	0.723
Miles to CH	No	31.749	63.006	62.286	Yes	31.749	30.911	40.576
Min dist CH-gram	Yes	1.932	1.514	1.014	No	1.932	1.752	2.336
1+ KS2 TA > tests levels	Yes	0.162	0.163	0.208	Yes	0.162	0.189	0.198
N		377	337	6,084		377	339	6,844

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

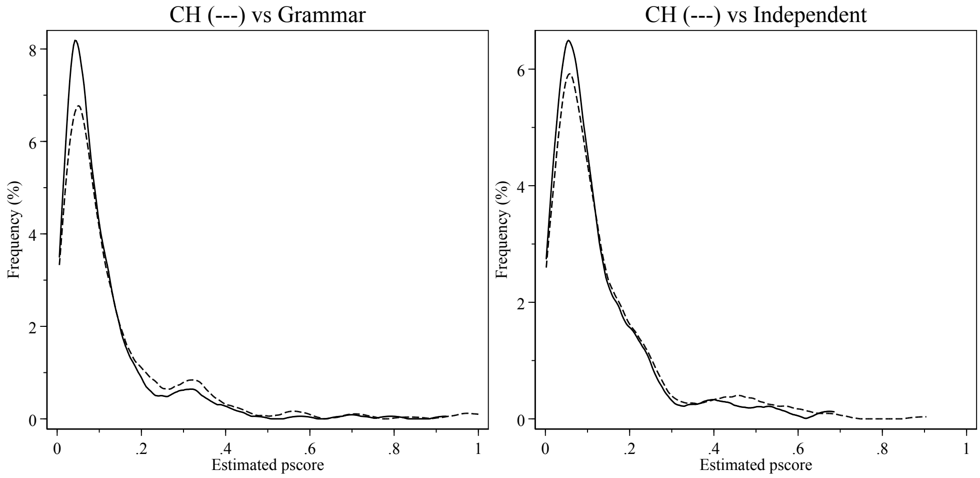
characteristics that were used to estimate the pscore. For example the dummy that equals 1 if a pupil's IDACI is greater than the median value of the index is used as a measure of SES when the control group are pupils in grammar schools, instead the IDACI score is used in the matching when the control group are pupils in independent schools. The table shows that differences between mean values for treated and matched controls are substantially smaller than the difference between mean values for treated and non-matched controls. This is the case for example for KS1 test achievement and for the distance in miles between a pupil's primary school and CH.[17,18]

Figure 2 shows the estimated propensity score distribution for CH pupils and for matched pupils in each of the two control groups separately. Reassuringly the relative frequency of low pscore values is higher for pupils in the control groups and vice versa for high pscore values. The *common support*, measured on the horizontal axis, is the interval of propensity score values over which the probability of observing pupils, measured on the vertical axis, is positive both for the control and for the

[17] Figure A.3 in the Appendix reports differences in predetermined characteristics separately for each value of the 8 estimated pscore blocks used in our specification, with the number of blocks determined by the matching algorithm (Becker and Ichino, 2002). Differences between treated and matched controls are small and not significant for each predetermined characteristic and in each pscore block.

[18] Figure A.4 in the Appendix shows scatterplots of the percentage of pupils who are eligible for FSM, as well as having an IDACI above the median, and the percentage of pupils who obtained the top level in Key Stage 2 tests, i.e. 5, in all three subjects, by using school-level data. CH stands out with an intake of low SES and high ability pupils.

Figure 2: Kernel density estimate of the propensity score



treatment group. This varies from 0 to about 0.9 and to 0.7 for grammar and independent schools respectively.

5.2 The effect of CH on achievement

In this section we report ATT estimates of the impact on achievement of attending a boarding selective school rather than a day selective school. Overall the positive and significant ATT estimates in Table 5 support our hypothesis that providing an academically oriented environment in non-school hours, such as the one in CH, improves attainment.

Table 5: Effect of attending CH on results in school-leaving exams

	Grammar			Independent		
	ATT	Controls		ATT	Controls	
		Matched	Non-matched		Matched	Non-matched
1+ GCSE with A	0.069**	0.867	0.865	0.093***	0.844	0.752
S.e.	0.023			0.024		
1+ GCSE with A*	0.172***	0.666	0.560	0.127**	0.711	0.518
S.e.	0.034			0.032		
5+ GCSE with A-A*	0.186***	0.584	0.505	0.172***	0.597	0.413
S.e.	0.036			0.035		
N		337	6,084		339	6,844

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We use nearest neighbour matching with replacement as our preferred method to match controls to treated and we set at 0.01 the maximum distance in pscore that is allowed to perform a match. The estimates in Table 5 are based on pupils within the common support and show that the probability of achieving at least 1 (1+ hereafter) GCSEs with A for CH pupils is significantly higher, by 6.4 and 9 percentage points (ppt hereafter) relative to grammar and independent schools respectively. This is 7.3% and 10.6% higher relative to the value for matched controls, also reported in Table 5. The probability of obtaining 1+ GCSEs with A* is also significantly higher, by 17 and 13.3 ppt or 25.5% and 18.8% relative to grammar and independent schools. Finally, the probability of obtaining 5+ GCSEs with A-A* is 18.6 and 18 ppt higher or 31.8% and 30.6% relative to the control groups respectively. Overall, the point estimates are higher when using the dummy equal to 1 if pupils obtain 1+ GCSEs with A* or 5+ GCSEs with A-A*. Pupils who obtain these results are approximately in the top two deciles of the distribution of achievement in GCSE exams among all pupils in the administrative data and in the top half if we only consider pupils in our control groups, as shown in Figure A.2 in the Appendix.[19]

In addition to ATT estimates and mean values of outcomes for matched controls, Table 5 shows mean values for non-matched controls. This allows us to compare our ATT estimates with naive estimates obtained as the difference in mean achievement between CH pupils and non-matched pupils in grammar and in independent schools. Naive estimates have the same sign as our ATT estimates although they are greater since the mean value of the outcomes for non-matched pupils is smaller than for matched ones. Under our untestable identifying assumption of selection on observables, naive estimates are then biased upwards relative to our ATT estimates. This comparison suggests that had pupils gone to grammar or independent day schools instead of CH, they would have obtained higher scores than the average in those schools.[20]

[19]Estimates obtained after excluding pupils outside the 'thick' support region of the pscore, which is defined in Black and Smith (2004) as a more conservative condition than the common support, are in line with our main results and can be found in Table A.1 in the Appendix. Our main results are also robust to allowing for multiple treatments, i.e. CH, grammar or independent schools, rather than a binary one, following Lechner (2002) and can be found in Table A.2 in the Appendix. In addition, results obtained using different matching methods, e.g. kernel or radius matching, are in line with our main results and can be found in Table A.3 in the Appendix. Finally, we find no difference in the CH effect by SES or gender. These results can be found in Table A.4 in the Appendix.

[20]Table A.5 in the Appendix reports estimates of the CH effect obtained using pscore estimates, as in Table 5 and OLS estimates. The table shows that under our identification assumption OLS estimates tend to differ from pscore estimates. Differences between the two sets of estimates are likely to be driven by the failure of the common support condition and the linear functional form assumption. Our main results are robust to using analytical standard errors and to a large-sample bias when matching on continuous variables, e.g. Key Stage 2 scores and the IDACI, by using

To summarise, ATT estimates of the effect of attending CH are positive and significant. They are greater when using proxies for high achievement at GCSE as outcomes (i.e. 1+ GCSEs at A* or 5+ GCSEs with A-A*) which is equivalent to the top end of the distribution of achievement at GCSE computed over pupils in all schools and in the top half when considering pupils in CH, grammar or independent schools. They are a robust to a number of changes in our preferred specification, as shown by additional results which can be found in Table A.1-A.3 in the Appendix.[21]

6 Sensitivity analysis

In this section we assess the validity of the conditional independence assumption (CIA), because it is the identifying assumption in our empirical analysis. In section 6.1 we report results of a placebo test on predetermined characteristics that were not included as controls in our pscore. In section 6.2 we implement the sensitivity analysis proposed by Ichino *et al.* (2008) to assess whether and to what extent our main results are affected by a failure of the CIA.

6.1 Placebo test

The administrative data we use in the empirical analysis include a rich set of pupil predetermined characteristics. To estimate the pscore we conservatively select from this set only those characteristics which are balanced in each block of the estimated pscore. The blocks are determined by the matching algorithm in Becker and Ichino (2002). This allows us to perform a placebo test to support the credibility of the conditional independence assumption (CIA) in our setting. Specifically we test whether the pre-treatment characteristics not included in the pscore estimation are balanced in the treatment and control groups. These variables which are defined

the methodology proposed in Abadie and Imbens (2006, 2011). We do not report these estimates although they are available upon request.

[21] We find that the balancing property is satisfied only when we use grammar school pupils as controls if we only match pupils based on a subset of all data available on achievement in Key Stage 2 tests, either on achievement levels or point scores. We find that the estimates of the CH effect relative to grammar school pupils tend to be 30% smaller for the probability of obtaining 1+ GCSE at A and, in contrast, 10-20% larger for the probability of obtaining 1+ GCSE at A-A* or 5+ GCSE at A-A*. When we use independent school pupils as controls the estimates of the CH effect cannot be interpreted meaningfully because the balancing property is not satisfied. We obtain similar results if we define all pupils in grammar or independent schools as our control group, rather than just those who were in primary schools in the same LAs as CH pupils. These results are not reported but are available upon request.

pseudo-outcomes in Imbens and Rubin (2015) are known to be unaffected by the treatment assignment under the CIA and therefore there should be no systematic differences in them by treatment status if the matching is successful.

Table 6: Difference in pre-treatments not used to match CH pupils with controls

	Grammar			Independent		
	ATT	Controls		ATT	Controls	
		matched	non-matched		matched	non-matched
KS1 Eng. reading DfE	.	0.070	-0.014	0.038	0.487	0.003
S.e.	.			0.055		
KS1 Eng. writing DfE	.	0.168	-0.021	.	0.537	-0.009
S.e.	.			.		
KS1 Mat. DfE	0.049	0.068	-0.017	.	0.486	-0.000
S.e.	0.078			.		
KS2 Eng TA ≥ 4	0.003	0.995	0.997	0.003	0.995	0.929
S.e.	0.005			0.005		
KS2 Mat TA ≥ 4	0.000	1.000	1.000	0.000	1.000	0.929
S.e.	0.000			0.000		
KS2 Sci TA ≥ 4	0.000	1.000	0.999	0.000	1.000	0.960
S.e.	0.000			0.000		
KS2 Eng TA > 4	-0.003	0.836	0.756	0.032	0.801	0.531
S.e.	0.030			0.030		
KS2 Mat TA > 4	-0.003	0.902	0.865	0.045	0.854	0.608
S.e.	0.023			0.026		
KS2 Sci TA > 4	0.032	0.875	0.838	0.050**	0.857	0.636
S.e.	0.025			0.025		
KS2 Eng TA $>$ test	-0.034	0.098	0.111	-0.027	0.090	0.095
S.e.	0.021			0.021		
KS2 Mat TA $>$ test.	0.016	0.058	0.072	-0.016	0.090	0.083
S.e.	0.020			0.021		
KS2 Sci TA $>$ test.	0.003	0.040	0.059	-0.005	0.048	0.054
S.e.	0.015			0.016		
KS2 Eng TA $\neq$ test	-0.029	0.172	0.211	-0.077**	0.220	0.228
S.e.	0.029			0.030		
KS2 Mat TA $\neq$ test.	0.011	0.114	0.126	-0.034	0.159	0.164
S.e.	0.026			0.026		
KS2 Sci TA $\neq$ test.	-0.016	0.117	0.150	-0.029	0.130	0.205
S.e.	0.025			0.025		
Miles to CH	-31.752***	63.501	62.286	0.108	31.641	40.576
S.e.	3.372			2.674		
Miles to closest grammar	0.042	1.889	1.014	0.242	1.689	2.336
S.e.	0.225			0.229		
Miles to closest indep.	-1.554***	2.568	2.263	0.265***	0.748	0.919
S.e.	0.224			0.097		
N		337	6,084		339	6,844

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The following pseudo-outcomes are used in our placebo test: Key Stage 1 achievement levels in Reading, Writing and Maths, rescaled following the approach proposed by the Department for Education (2017); a set of binary outcomes based on pupil teacher assessment (TA) at the end of Key Stage 2; a set of proxies for pupil non-cognitive skills consisting of dummies equal to 1 if at Key Stage 2 the difference between a pupil achievement level in the TA and in tests by subject is positive or negative. Additional information about our proxies for non-cognitive skills can be found in section 3.4. Finally we include measures of distance in miles between a pupil's primary school, CH and the closest grammar or independent schools.

Table 6 reports pscore matching estimates of the difference in the aforementioned pre-treatment characteristics for CH pupils relative to their matched controls in grammar and in independent schools. The differences are small or close to zero and not significant for all measures of achievement at Key Stage 1, for all measures of TA and of difference between TA and test achievement at Key Stage 2 when the control group are pupils in grammar school and for all but two when the control group are pupils in independent schools. Two out of three measures of distance from primary school are balanced when the control group are pupils in independent schools. Only one out of three is balanced when the controls are grammar school pupils. This is due to the considerably lower number of grammar relative to independent schools mechanically leading to lower distance measure for CH pupils relative to matched controls in grammar schools, as shown in Figure 1. Overall, these results suggest that the CIA is plausible in our empirical setting.[22]

Finally, we performed an additional placebo test with the same logic as the aforementioned test, except we now quantify the differences in Key Stage 2 tests for CH pupils and matched pupils in grammar and independent schools by conditioning only on their achievement in Key Stage 1 tests and on socio-demographics. The aim of this test is to assess whether it is possible to find control students whose achievement predicts the future achievement of CH before the start of secondary school. Table

[22]The missing values in Table 6 indicate that a given predetermined characteristic was used for a given control group and therefore the placebo test with that predetermined characteristic was not performed as it would not be informative. Extra results of our placebo test obtained using other measures of achievement at Key Stage 1 can be found in Table A.6 in the Appendix. We also performed a slightly different version of the placebo test in which we only estimate the propensity score by using achievement at Key Stage 2 and socio-demographics. Thus we fully exclude from the pscore estimation all measures of achievement at Key Stage 1 and we assess whether these are still balanced. Table A.7 in the Appendix reports the results. Differences in measures of achievement in Key Stage 1 tests tend to be balanced for both control groups, except for reading.

A.8 in the Appendix reports pscore matching estimates obtained by adding dummies for whether achievement by subject in Key Stage 1 tests was above the median to the main measures of achievement in Key Stage 1 (see Table 4). This pscore specification passes conventional diagnostics as covariates are balanced in each pscore block while they are not if we add to our main specification the additional measures of achievement in Key Stage 1 tests. Table A.8 shows that Key Stage 2 achievement by subject is balanced for CH pupils and for matched controls in grammar schools with similar Key Stage 1 achievement and socio-demographics except for English and science where CH pupils have higher achievement. In contrast, achievement at Key Stage 2 is only slightly balanced when matched controls are pupils in independent schools, with significantly higher achievement of CH pupils in all subjects. Overall, this placebo tests confirms that CH pupils and matched grammar schools pupils are overall comparable while independent school pupils less so, which is not a complete surprise since independent schools tend to be less selective on merit relative to grammar schools, as we discussed in section 3.3.

However this placebo test has a number of limitations due to the low performance of Key Stage 1 assessment as a predictor of achievement in Key Stage 2 tests. While CH pupils are all high achievers in Key Stage 2, not all high achievers in Key Stage 2 perform equally well in their previous assessment (Key Stage 1). This is shown in Table A.9 in the Appendix by the 20-30% of pupils with low performance in Key Stage 1 and high performance in Key Stage 2 tests both when we use data on all pupils in our dataset and when we focus only on CH pupils and pupils matched to them based on similar values of the propensity score used in our main specification. First, Key Stage 1 tests are not externally assessed, while Key Stage 2 tests are, which may introduce a bias due to teachers' unobservable characteristics. Secondly, data on Key Stage 1 achievement is less precisely measured since pupils obtain categorical information about the level achieved in each subject while in Key Stage 2 tests they obtain both integer scores in the interval 1-100 and the corresponding categorical achievement level. Thirdly, pupils take Key Stage 1 tests at age 7. This is an early age at which pupils exhibit high heterogeneity in aspects of cognitive development, such as cognitive flexibility, goal setting and information processing, which are positively associated with educational achievement. By the age of 9, empirical research in psychology shows that heterogeneity in cognitive development decreases (Anderson, 2002).

6.2 Sensitivity of CH effect to calibrated confounders

We assess the sensitivity of our main results to a failure of the CIA by implementing the methodology proposed by Ichino *et al.* (2008). This consists of considering a binary unobservable confounder that has an effect on the both the selection into treatment and on the untreated outcome. An example of unobservable confounders in our setting are non-cognitive skills since they can have an effect on the probability of attending CH and an effect on GCSE achievement, which can differ for pupils in CH relative to those in grammar or independent schools. By imposing the parameters of the confounder distribution we can predict a value of the confounder for each individual in our sample. The simulated confounder can then be added to the set of matching variables to obtain an estimate of the ATT. This procedure is repeated 1000 times and the final estimated ATT is the average of the ATTs over the distribution of the simulated confounder. If this ATT, obtained including the simulated confounder, is similar to the ATT obtained without confounder, the CIA is more likely to hold than if they are substantially different.

Similarly, we can use the simulated values of the unobservable confounder to obtain an estimate of its effect on the relative probability of a positive outcome for the non-treated (outcome effect) and on the relative probability of treatment (selection effect). These relative probabilities are obtained as average odds ratios after estimating for each set of simulated confounder values a logit model for the probability of a positive outcome of the untreated and another one for the probability of treatment. We now provide a more detailed description of this methodology.

We let U be the unobserved binary confounder and its distribution be fully determined by the four parameters $p_{ij} = Pr(U = 1 \mid D = i, A = j, X)$ measuring the probability that the unobserved term is equal to 1 given that the treatment D, i.e. attending CH in our setting, is equal to i and the outcome A, i.e. achievement, is equal to j, with $i, j = \{0, 1\}$.

$$\Gamma = \frac{\frac{Pr(A=1 \mid D=0, U=1, X)}{Pr(A=0 \mid D=0, U=1, X)}}{\frac{Pr(A=1 \mid D=0, U=0, X)}{Pr(A=0 \mid D=0, U=0, X)}} \tag{2}$$

By assuming $p_{01} > p_{00}$, i.e. that the unobserved confounder has a positive effect on the untreated outcome, and accounting for the relationship between U and X, Ichino *et al.* (2008) define the outcome effect Γ as the effect of U on the probability of a positive outcome A and compute it as the odds ratio of U after estimating the logit model of $Pr(A = 1 \mid D = 0, U, X)$, as shown in equation (2). The selection effect Δ is defined as the effect of U on the probability of treatment, i.e. $D = 1$, and it is computed as the odds ratio of U after estimating the logit model of $Pr(D = 1 \mid U, X)$, as shown in equation (3).

$$\Delta = \frac{\dfrac{Pr(D=1 \mid U=1, X)}{Pr(D=0 \mid U=1, X)}}{\dfrac{Pr(D=1 \mid U=0, X)}{Pr(D=0 \mid U=0, X)}} \tag{3}$$

Based on values of p_{ij} obtained by using the empirical distribution of a relevant covariate, a value of U is imputed for each pupil in the dataset. The variable U is then treated as any observed covariate in X to first estimate the pscore and then the ATT using nearest neighbour matching. Varying the values of the sensitivity parameters p_{ij} and repeating the pscore and ATT estimation in a simulation 1000 times, the average of the ATT over the distribution of U is obtained.[23]

In our setting, achievement in Key Stage 2 tests at age 11 and SES are observable characteristics used by CH to select its pupils while suitability for boarding is unobservable to the econometrician. Therefore, we assess the sensitivity of our main results to unobserved binary covariates whose distribution is similar to the one of observed measures of pupil ability, as at least part of a pupil's ability is typically unobserved and may be correlated with the pupil's resilience to adapt to boarding.

We use two measures of achievement in Key Stage 1 test as ability proxies, the English comprehension test and the Maths task test, rescaled following Department for Education (2017). Since we do not use them as predetermined characteristics when we estimate the pscore they may differ for pupils in the treated and control groups. We also use a dummy equal to 1 if the distance between primary school and the closest grammar or independent secondary is greater than the median value, because it is an observable measure of the opportunity cost of attending CH although is not used in estimating the pscore. This may be a relevant factor for secondary school choice as the further away a pupil lives from CH the higher the psychological effort required to adapt to boarding.

Panel A of Table 7 shows estimates of the effect of CH obtained on our three measures of achievement at GCSE by using pupils in grammar schools as controls. Estimates on each row are obtained by using a confounder U distributed accord-

[23] A more detailed description of the econometric details behind the sensitivity analysis is found in section 4 in Ichino *et al.* (2008).

Table 7: Sensitivity analysis of CH effect using confounders calibrated following ability and distance proxies

	p_{11}	p_{10}	p_{01}	p_{00}	Outcome effect Γ	Selection effect Δ	ATT	S.e.
Panel A: grammar schools								
1+ GCSEs with A								
Neutral confounder	0.515	0.517	0.493	0.485	1.031	1.112	0.072***	0.024
KS1 Eng compr. test DfE	0.005	0.107	0.005	0.006	1.200	2.387	0.049*	0.027
KS1 Mat task test DfE	0.008	0.069	0.001	0.000	.	12.497	0.045	0.027
Closest grammar school	0.530	0.517	0.496	0.498	0.996	1.149	0.048**	0.027
*1+ GCSEs with A**								
Neutral confounder	0.501	0.553	0.499	0.496	1.014	1.058	0.172***	0.034
KS1 Eng compr. test DfE	0.015	0.000	0.009	0.000	.	2.276	0.185***	0.041
KS1 Mat task test DfE	0.015	0.000	0.002	0.000	4.101	12.752	0.188***	0.041
Closest grammar school	0.521	0.566	0.487	0.508	0.919	1.152	0.193***	0.042
*5+ GCSEs with A-A**								
KS1 Eng compr. test DfE	0.013	0.011	0.010	0.000	27.628	2.279	0.187***	0.044
KS1 Mat task test DfE	0.016	0.000	0.002	0.000	.	12.064	0.190***	0.044
Closest grammar school	0.517	0.570	0.483	0.510	0.901	1.158	0.188***	0.045
Panel B: independent schools								
1+ GCSEs with A								
Neutral confounder	0.480	0.586	0.497	0.513	0.938	0.953	0.093***	0.024
KS1 Eng compr. test DfE	0.005	0.107	0.006	0.006	1.074	2.324	0.105***	0.031
KS1 Mat. task test DfE	0.708	0.655	0.548	0.276	3.154	2.395	0.071***	0.030
Closest indep. school	0.470	0.414	0.498	0.514	0.937	0.878	0.103***	0.031
*1+ GCSEs with A**								
Neutral confounder	0.524	0.474	0.500	0.490	1.044	1.090	0.127***	0.032
KS1 Eng compr. test DfE	0.015	0.000	0.009	0.001	12.526	1.866	0.142***	0.038
KS1 Mat task test DfE	0.739	0.536	0.655	0.287	4.746	1.988	0.095**	0.040
	0.374	0.447	0.492	0.521	0.893	0.627	0.130***	0.040
Closest indep. school	0.467	0.461	0.472	0.535	0.779	0.914	0.136***	0.040
*5+ GCSEs with A-A**								
Neutral confounder	0.483	0.450	0.497	0.494	1.013	0.925	0.172***	0.035
KS1 Eng compr. test DfE	0.013	0.011	0.011	0.002	13.169	1.876	0.164***	0.042
KS1 Mat task test DfE	0.775	0.474	0.701	0.320	4.961	1.933	0.117***	0.044
Closest indep. school	0.468	0.460	0.473	0.523	0.818	0.912	0.158***	0.043

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

ing to the aforementioned predetermined characteristics not used to estimate the propensity score. The first four columns on the left-hand side show values of the probabilities p_{ij} characterising the distribution of U by using the empirical distribution of the covariate on a given row, then the outcome and selection effect are shown and, finally, ATT estimates and their standard error. Panel B shows the the same information when pupils in independent schools are the control group.

For each outcome variable, Table 7 shows estimates obtained using a neutral confounder, i.e. with all p_{ij} set equal to approximately 0.5, in the first row. On the two following rows the unobserved confounder is distributed similarly to observed measures of achievement at Key Stage 1. In the final row the confounder follows the empirical distribution of a dummy equal to 1 if the distance to the closest grammar secondary or to the closest independent secondary is greater than the median.

Overall, the estimates in Table 7 show that both their magnitude and precision are in line with our main results. When we look instead at the outcome effect, i.e. the effect of U on the probability of higher achievement, and at the selection effect, i.e. the effect of U on the probability of attending CH, the table shows that the value of both effects is very close to one in the case of neutral confounder. This is expected as by setting all p_{ij} to 0.5 the confounder is close to i.i.d. If we now focus on proxies for unobserved ability, the majority of the values of the outcome and selection effect are greater than 1, with the outcome effect being greater. This suggests a positive selection into CH and a positive effect on achievement for CH pupils with high unobserved ability.

For distance measures, the results in Table 7 show that both the outcome and selection effect of a confounder proxying the cost of attending a selective day school are close to 1. The selection effect tends to be greater than one when pupils in grammar schools are the control group, which suggests a positive selection effect due to a lower opportunity cost of attending CH if the closest selective day school is far. When pupils in independent schools are the control group, the selection effect is slightly smaller but very close to one, suggesting that the selection effect plays a small role. These results hold qualitatively for all the three outcomes we consider. Additional results using measures of socio-demographics, other measures of achievement at Key Stage 1, the type of school attended at Key Stage 2 and measures of motivation not used in the pscore estimation can be found in Table A.10 in the Appendix.[24]

7 Discussion

In this paper we use English administrative data to test the hypothesis that attending Christ Hospital (CH), a boarding school admitting a higher share of high ability

[24]The analysis of killer confounders in Ichino *et al.* (2008), which consists of jointly increasing the extent of selection and of outcome effects until a pair of values for these effects that 'kills' the main results is found, is not reported as it is not informative about the nature of the unobserved information, e.g. ability or opportunity cost, that may bias our main results.

pupils with lower socio-economic status (SES) than comparable selective day schools improves achievement in compulsory school final exams (GCSEs). Our propensity score matching estimates show that the probability of achieving A or A* in five or more GCSEs is 18.6 percentage points higher compared to 58.4% for matched pupils in grammar schools, i.e. an increase by about 30%. We find similar results when the control group are independent school pupils.

CH differs from independent day schools in that it provides boarding and tends to be more selective based on pupil ability. Therefore, when pupils in independent day schools are the control group our estimates can be interpreted as the joint effect of boarding and of pupil ability selection. However, we argue that the potential selection bias in our estimates is small. The reason is that since independent schools display a high variability in pupil ability, ranging from very high for pupils admitted with a bursary to a lower level for fee-paying pupils, high ability pupils at independent schools can be repeatedly matched to similar pupils at CH, thus making boarding the most plausible mechanism underlying our estimated effect.

On the other hand when pupils in grammar schools are the controls, our estimates capture the overall effect of substituting family with school inputs and of having access to better school inputs since CH is boarding and has more resources. Although we cannot separately quantify the boarding effect and the resources effect without additional assumptions, the fact that we obtain similar results with the independent day schools control group, where resources are much closer to those of CH, suggests that boarding is an important part of the explanation for the difference between exams performances of matched pupils at CH and in grammar school.

Our results contribute to recent studies exploiting lottery-based admission into oversubscribed boarding schools, in the US (Curto and Fryer Jr, 2014) and in France (Behaghel *et al.*, 2017), by being the first to undertake a comparable analysis on England. By estimating the treatment effect on the treated (ATT), we offer complementary evidence to the quasi-experimental one obtained using a local average treatment effect (LATE). A common limitation of our study and the two related quasi-experimental ones on boarding schools is low external validity, as they all use either a single boarding school or a small number of them as the treatment group, which makes them unrepresentative of the universe of boarding schools in a country.

ATT has a somewhat "stronger" identification assumption based on selection on observables, i.e. lower internal validity relative to quasi-experimental studies. While we would like to match pupils by restricting our attention to those applying to CH and compare achievement at GCSE for marginally (non-)admitted ones based on their performance in the selection process, this is not possible due to privacy laws. However, our extensive sensitivity analysis that includes a placebo test assessing the differences in predetermined characteristics not used to match pupils and a methodology developed by Ichino *et al.* (2008) to test the extent to which selection on unobservables may bias our estimates, offers support to our estimation strategy based on selection on observables. In our study, this suggests that unobservables associated with admission to CH such as ability, motivation or the opportunity cost of attending a different school, do not play a major role in our analysis. We believe that this is consistent with the fact that both the treated group, i.e. CH pupils, and the control group, i.e. either grammar or independent school pupils, are screened based on ability and motivation.

Our paper also contributes to empirical studies estimating an educational production function to assess the effect of school-based policies that aim to counteract the negative influence of low SES on pupil achievement (see for a survey Todd and Wolpin, 2003). We isolate the effect of boarding in a simple setting in which parental inputs, which can be a confounding factor of shool-based interventions, are low for all boarders. This cannot be done in the production function, because unobserved family inputs may either decrease if school and family inputs are substitutes or increase if they are complements.

Our analysis paves the way for a number of extensions. We have so far focused on a single selective and boarding school but also considering state boarding schools, a number of which are Academies, may help us obtaining as treatment group one that is more representative of secondary school pupils than the highly selected one at CH. We have not yet looked at the probability of continuing with post-compulsory education, namely sixth form, achievement in A-levels, admission into prestigious universities, degree choice and achievement and labour market outcomes since access to this individual-level data is subject to authorisation by the Department for Education. In addition, an extension that is particularly relevant to inform policy-decisions over the role of boarding education for high ability pupils with low SES is performing a cost-benefit analysis of subsidising these pupils. Finally, peer effects are assumed away under all reduced form empirical strategies, such as propensity score matching or instrumental variables. However, they may play a role for CH pupils and for those in selective secondary day schools. While we believe that this is an interesting and partly unanswered question, the lack of data on pupils' networks limits the possibility of exploring this empirical issue.

References

Abadie, A. and Imbens, G. W. (2006). Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica*, **74**, 235–267.

— and — (2011). Bias-corrected Matching Estimators for Average Treatment Effects. *Journal of Business and Economic Statistics*, **29** (1), 1–11.

Allen, R. and Bartley, J. (2017). The role of the eleven-plus test papers and appeals in producing social inequalities in access to grammar schools. *National Institute Economic Review*, **240** (1), R30–R41.

Anderson, P. (2002). Assessment and Development of Executive Function (EF) During Childhood. *Child Neuropsychology*, **8** (2), 71–82.

Andersson, C. and Johansson, P. (2013). Social Stratification and Out-of-School Learning. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **176** (3), 679–701.

Becker, S. O. and Ichino, A. (2002). Estimation of Average Treatment Effects Based on Propensity Scores. *The Stata Journal*, **2** (4), 358–377.

Behaghel, L., de Chaisemartin, C. and Gurgand, M. (2017). Ready for Boarding? The Effects of a Boarding School for Disadvantaged Students. *American Economic Journal: Applied Economics*, **9** (1), 140–164.

Black, D. A. and Smith, J. A. (2004). How Robust is the Evidence on The Effects of College Quality? Evidence from Matching. *Journal of Econometrics*, **121** (1), 99–124.

Burgess, S. and Greaves, E. (2013). Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities. *Journal of Labor Economics*, **31** (3), 535–576.

—, —, Vignoles, A. and Wilson, D. (2015). What Parents Want: School Preferences and School Choice. *The Economic Journal*, **125** (587), 1262–1289.

Chowdry, H., Crawford, C., Dearden, L., Goodman, A. and Vignoles, A. (2013). Widening participation in higher education: analysis using linked administrative data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **176** (2), 431–457.

Crawford, C. and Greaves, E. (2013). *A Comparison of Commonly Used Socioeconomic Indicators: their Relationship to Educational Disadvantage and Relevance to Teach First*. Ifs reports, Institute for Fiscal Studies.

—, Macmillan, L. and Vignoles, A. (2014). Progress Made by High-attaining Children from Disadvantaged Backgrounds. *Social Mobility and Child Poverty Commission (Available from https://goo.gl/0emqZ8f)*.

Curto, V. E. and Fryer Jr, R. G. (2014). The Potential of Urban Boarding Schools for the Poor: Evidence from SEED. *Journal of Labor Economics*, **32** (1), 65–93.

Dearden, L., Sibieta, L. and Sylva, K. (2011). The Socio-economic Gradient in Early Child Outcomes: Evidence from the Millennium Cohort Study. *Longitudinal and Life Course Studies*, **2** (1), 19–40.

Department for Education (2014). Boarding Schools. An Opportunity to Improve Outcomes for Vulnerable Children. *Department for Education (Available from https://goo.gl/T0rosq)*, (DFE-00669-2014).

Department for Education (2016). Schools That Work for Everyone. Government Consultation. *Department for Education (Available from https://goo.gl/uYjmFc)*, (DFE-00146-2018).

Department for Education (2017). Attainment in Primary Schools in England. *Department for Education (Available from https://goo.gl/KxzLte)*, (SFR 49/2017).

Epple, D., Romano, R. and Zimmer, R. (2016). Charter Schools: A Survey of Research on Their Characteristics and Effectiveness. In S. M. Eric A. Hanushek and L. Woessmann (eds.), *Handbook of the Economics of Education*, vol. 5, *3*, Elsevier, pp. 139 – 208.

Eyles, A., Hupkau, C. and Machin, S. (2016). School Reforms and Pupil Performance. *Labour Economics*, **41**, 9–19.

Fisher, S., Frazer, N. and Murray, K. (1986). Homesickness and Health in Boarding School Children. *Journal of Environmental Psychology*, **6** (1), 35–47.

Frölich, M., Heshmati, A. and Lechner, M. (2004). A Microeconometric Evaluation of Rehabilitation of Long-term Sickness in Sweden. *Journal of Applied Econometrics*, **19** (3), 375–396.

Gerfin, M. and Lechner, M. (2002). A Microeconometric Evaluation of the Active Labour Market Policy in Switzerland. *Economic Journal*, **112** (482), 854–893.

Green, F., Machin, S., Murphy, R. and Zhu, Y. (2012). The Changing Economic Advantage from Private Schools. *Economica*, **79** (316), 658–679.

Hodges, J., Sheffield, J. and Ralph, A. (2016). Staff and Boarders Perspectives of the Boarding Environment. *Journal of Child and Family Studies*, **25** (4), 1045–1056.

Ichino, A., Mealli, F. and Nannicini, T. (2008). From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and their Sensitivity? *Journal of Applied Econometrics*, **23** (3), 305–327.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An introduction*. Cambridge University Press.

Independent Schools Council (2014). *ISC Annual Census 2014 (Available from https://goo.gl/qktHh4)*. Tech. rep.

Jerrim, J. (2017). Comparing Socio-economic Gaps in the Performance of Highly Able UK Pupils Internationally. *Sutton Trust Report (Available from https://goo.gl/3nZ4Lw)*.

Lechner, M. (2002). Some Practical Issues in the Evaluation of Heterogeneous Labour Market Programmes by Matching Methods. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **165** (1), 59–82.

Lester, L., Mander, D. J. and Cross, D. (2015). Bullying Behaviour Following Students' Transition to a Secondary Boarding School Context. *International Journal of Child and Adolescent Health*, **8** (2), 141.

Leuven, E. and Sianesi, B. (2015). PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing. *Statistical Software Components*.

Machin, S., McNally, S. and Meghir, C. (2004). Improving Pupil Performance in English Secondary Schools: Excellence in Cities. *Journal of the European Economic Association*, **2** (2-3), 396–405.

—, — and Ruiz-Valenzuela, J. (2018). *Entry Through the Narrow Door: The Costs of Just Failing High Stakes Exams*. Research Discussion Paper 014, LSE Centre for Vocational Educational Research.

Martin, A. J., Papworth, B., Ginns, P. and Liem, G. A. D. (2014). Boarding School, Academic Motivation and Engagement, and Psychological Well-Being: A Large-Scale Investigation. *American Educational Research Journal*.

Reardon, S. F. (2011). The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations. *Whither opportunity*, pp. 91–116.

Schaverien, J. (2004). Boarding School: the Trauma of the Privileged Child. *Journal of Analytical Psychology*, **49** (5), 683–705.

— (2011). Boarding School Syndrome: Broken Attachments a Hidden Trauma. *British Journal of Psychotherapy*, **27** (2), 138–155.

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, **75** (3), 417–453.

Todd, P. E. and Wolpin, K. I. (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal*, **113** (485), F3–F33.

— and — (2007). The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital*, **1** (1), 91–136.

Watson, G. (2006). Computing Helmert transformations. *Journal of Computational and Applied Mathematics*, **197** (2), 387 – 394.

Wires, J. W., Barocas, R. and Hollenbeck, A. R. (1994). Determinants of Adolescent Identity Development: A Cross-sequential Study of Boarding School Boys. *Adolescence*, **29** (114), 361.

Online Appendix

Figure A.1: Number of CH pupils by local authority where they went to primary school

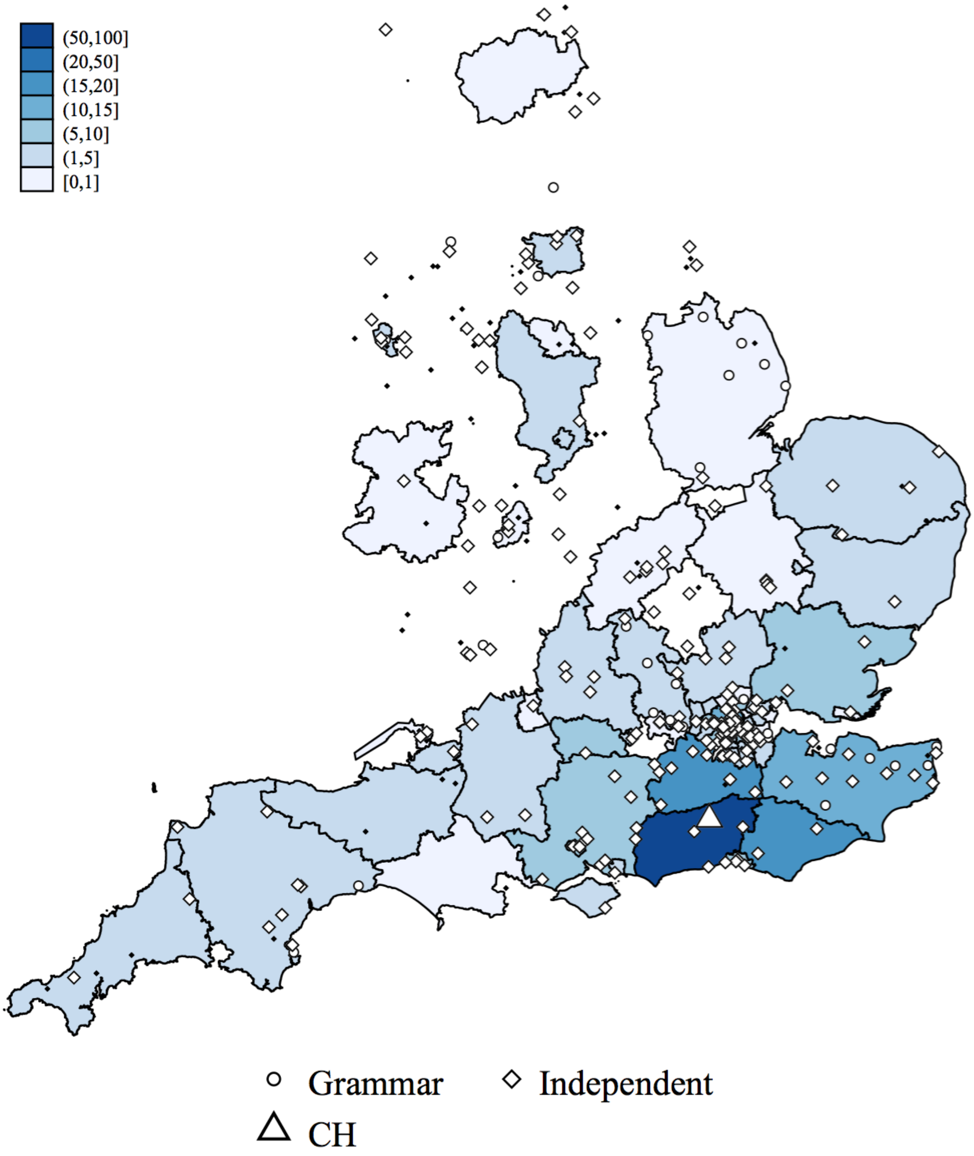
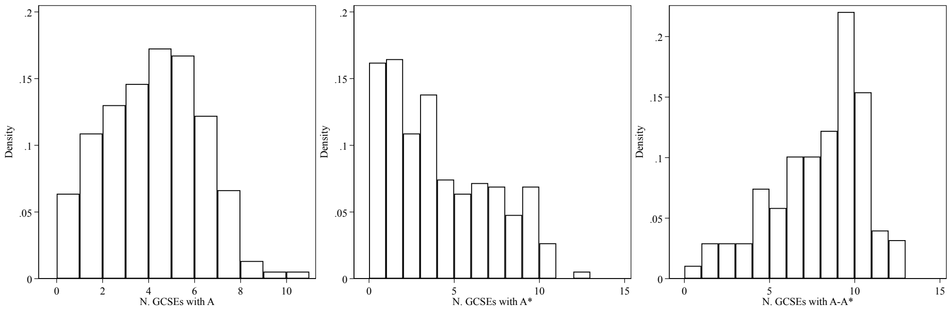
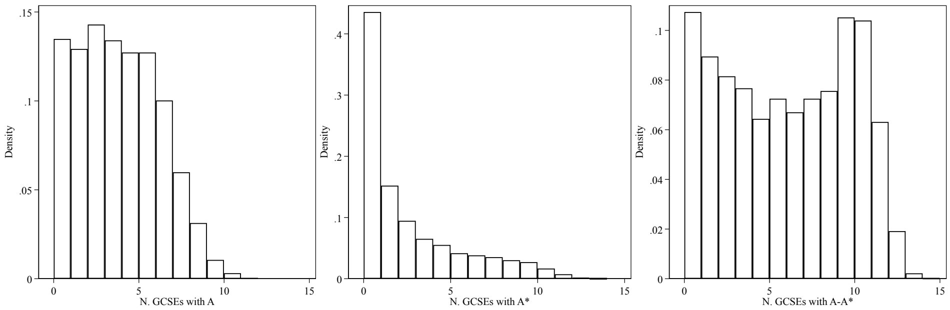


Figure A.2: Histograms of achievement at GCSE

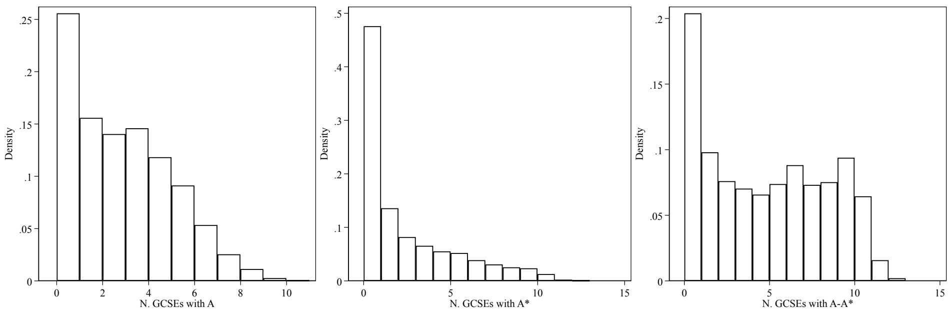
CH



Grammar



Independent



All schools

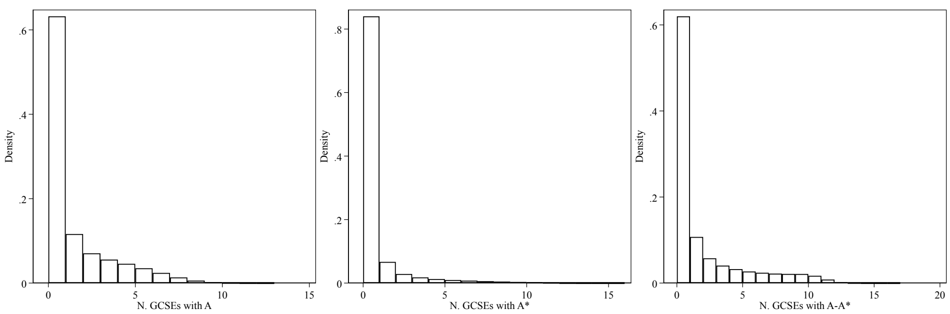


Figure A.3: Covariates differences for CH relative to grammar and independent by pscore block

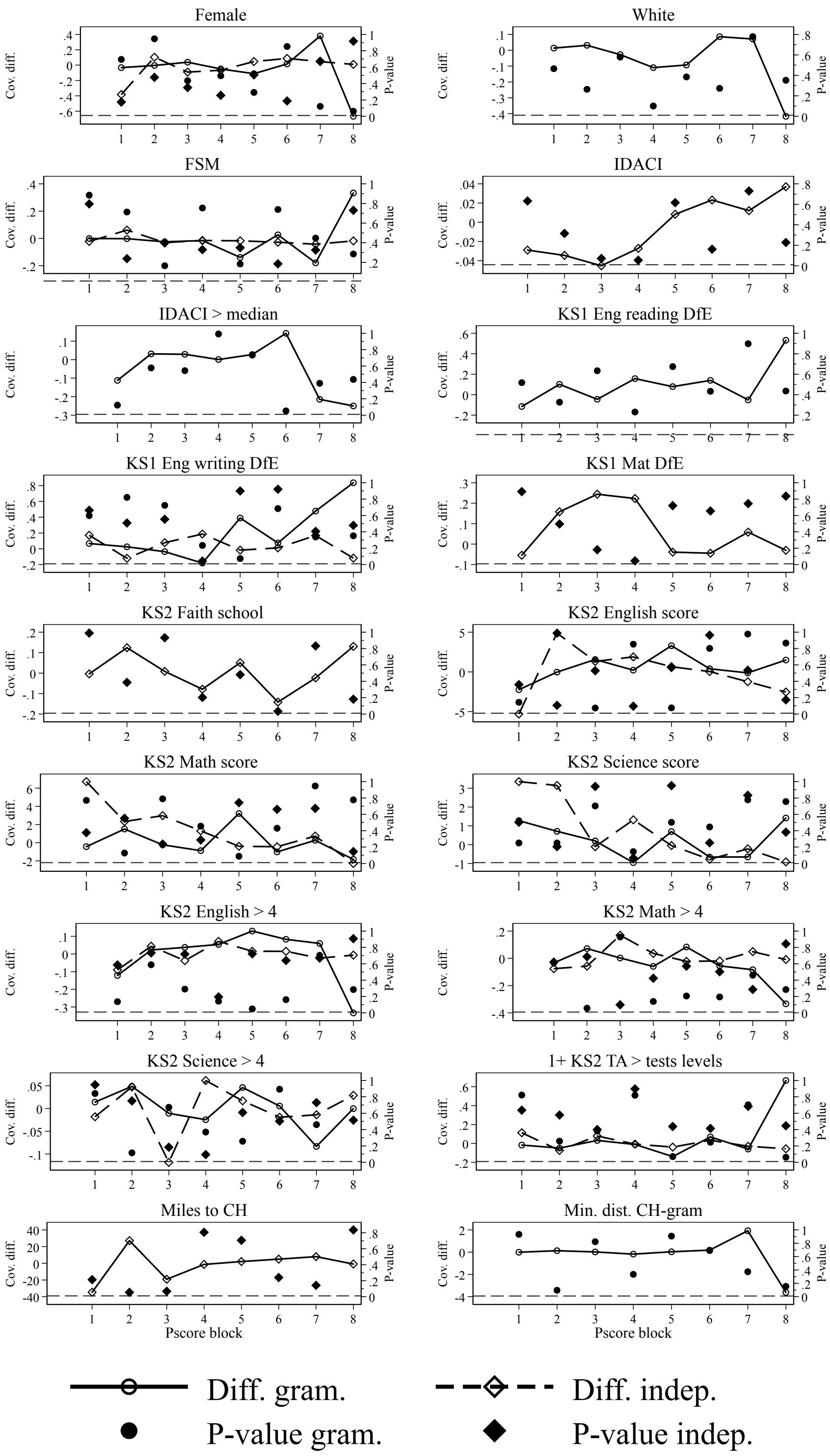


Figure A.4: Achievement at Key Stage 2 (KS2) and SES by secondary school

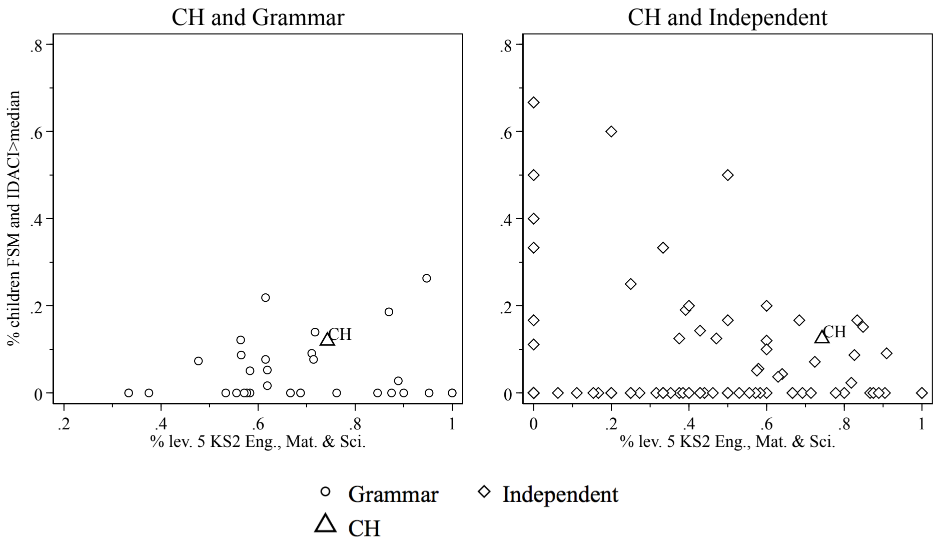


Table A.1: Effect of attending CH for pupils in the pscore thick support

	Grammar schools			Independent schools		
	ATT	Controls		ATT	Controls	
		Matched	Non-matched		Matched	Non-matched
			Pscore thick support 0-0.2			
1+ GCSE with A	0.078***	0.866	0.864	0.081***	0.856	0.752
S.e.	0.024			0.025		
1+ GCSE with A*	0.173***	0.657	0.558	0.084**	0.742	0.516
S.e.	0.036			0.035		
5+ GCSE with A-A*	0.193***	0.575	0.504	0.158***	0.611	0.411
S.e.	0.038			0.039		
N		293	6,023		279	6,763
			Pscore thick support 0-0.1			
1+ GCSE with A	0.088***	0.855	0.861	0.009	0.915	0.745
S.e.	0.029			0.027		
1+ GCSE with A*	0.185***	0.626	0.546	0.042	0.745	0.495
S.e.	0.042			0.042		
5+ GCSE with A-A*	0.189***	0.542	0.492	0.118**	0.590	0.386
S.e.	0.045			0.047		
N		223	5,534		204	6,280

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Estimates of CH effect using pscore from multinomial logit

	Grammar			Independent		
	ATT	Controls		ATT	Controls	
		Matched	Non-matched		Matched	Non-matched
1+ GCSE with A	0.051**	0.885	0.865	0.109***	0.828	0.753
S.e.	0.023			0.026		
1+ GCSE with A*	0.181***	0.656	0.530	0.154**	0.684	0.511
S.e.	0.035			0.033		
5+ GCSE with A-A*	0.176***	0.595	0.478	0.146**	0.623	0.391
S.e.	0.037			0.036		
N		331	1,595		331	2,774

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Estimates of CH effect obtained using different matching methods

	Grammar						
	Nearest neighbour		Kernel			Radius (0.1)	Mahalanobis
	Becker and Ichino (2002)	Leuven and Sianesi (2015)	Normal	Bivariate	Epanechnikov		
1+ GCSE with A	0.064**	0.072***	0.059***	0.051***	0.051***	0.059***	0.072***
S.e.	0.023	0.024	0.015	0.015	0.015	0.015	0.022
1+ GCSE with A*	0.170***	0.172***	0.231***	0.206***	0.209***	0.229***	0.172***
S.e.	0.034	0.034	0.022	0.023	0.023	0.022	0.029
5+ GCSE with A-A*	0.186***	0.188***	0.223***	0.203***	0.206***	0.222***	0.202***
S.e.	0.036	0.036	0.024	0.025	0.025	0.025	0.032
N	6,421	6,798	6,798	6,798	6,798	6,798	
	Independent						
	Nearest neighbour		Kernel			Radius (0.1)	Mahalanobis
	Becker and Ichino (2002)	Leuven and Sianesi (2015)	Normal	Bivariate	Epanechnikov		
1+ GCSE with A	0.090***	0.093***	0.145***	0.124***	0.128***	0.151***	0.042**
S.e.	0.024	0.024	0.015	0.016	0.016	0.015	0.021
1+ GCSE with A*	0.133***	0.127***	0.226***	0.166***	0.175***	0.229***	0.114***
S.e.	0.032	0.032	0.021	0.022	0.022	0.021	0.030
5+ GCSE with A-A*	0.180***	0.172***	0.253***	0.188***	0.198***	0.255***	0.119***
S.e.	0.036	0.035	0.024	0.025	0.024	0.024	0.033
N	7,183	7,560	7,560	7,560	7,560	7,560	7,560

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Split sample estimates of the CH effect

	Gender					
	Grammar			Independent		
	Male	Female	F-M	Male	Female	F-M
1+ GCSE with A	0.048	0.054	0.006	0.052	0.090***	0.037
S.e.	0.031	0.034	0.046	0.031	0.034	0.046
1+ GCSE with A*	0.190***	0.180***	-0.011	0.057	0.132***	0.075
S.e.	0.049	0.045	0.067	0.046	0.042	0.062
5+ GCSE with A-A*	0.186***	0.186***	-0.000	0.081	0.180***	0.099
S.e.	0.051	0.049	0.071	0.050	0.048	0.070
N	393	319		394	320	

	FSM					
	Grammar			Independent		
	No	Yes	Yes-No	No	Yes	Yes-No
1+ GCSE with A	0.037	0.073	0.035	0.090***	0.127*	0.037
S.e.	0.024	0.069	0.073	0.026	0.076	0.080
1+ GCSE with A*	0.127***	0.182*	0.054	0.065**	0.418***	0.353***
S.e.	0.035	0.102	0.108	0.033	0.105	0.110
5+ GCSE with A-A*	0.152***	0.145	-0.007	0.149***	0.327***	0.178
S.e.	0.038	0.109	0.116	0.037	0.110	0.117
N	620	93		620	91	

	IDACI greater than or equal to the median					
	Grammar			Independent		
	No	Yes	Yes-No	No	Yes	Yes-No
1+ GCSE with A	0.064	0.072***	0.007	0.035	0.122***	0.087
S.e.	0.040	0.028	0.049	0.040	0.030	0.050
1+ GCSE with A*	0.107*	0.215***	0.108	0.009	0.195***	0.186**
S.e.	0.056	0.041	0.069	0.059	0.039	0.071
5+ GCSE with A-A*	0.086	0.194***	0.108	0.052	0.229***	0.177**
S.e.	0.058	0.044	0.073	0.064	0.043	0.077
N	271	448		222	489	

	FSM and IDACI greater than or equal to the median					
	Grammar schools			Independent schools		
	No	Yes	Yes-No	No	Yes	Yes-No
1+ GCSE with A	0.039	0.067	0.028	0.082***	0.062	-0.020
S.e.	0.024	0.072	0.076	0.025	0.072	0.077
1+ GCSE with A*	0.130***	0.222**	0.093	0.073**	0.354***	0.281**
S.e.	0.035	0.105	0.111	0.033	0.109	0.114
5+ GCSE with A-A*	0.160***	0.156	-0.004	0.155***	0.250**	0.095
S.e.	0.038	0.115	0.121	0.037	0.117	0.123
N	636	79		633	81	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Continued on the next page

	FSM and IDACI in the top quartile					
	Grammar			Independent		
	No	Yes	Yes-No	No	Yes	Yes-No
1+ GCSE with A	0.050	0.051	0.001	0.084	0.119	0.035
S.e.	0.024	0.077	0.081	0.025	0.083	0.087
1+ GCSE with A*	0.148	0.231	0.083	0.069	0.452	0.384
S.e.	0.035	0.113	0.118	0.032	0.114	0.119
5+ GCSE with A-A*	0.172	0.154	-0.018	0.143	0.310	0.166
S.e.	0.037	0.124	0.129	0.036	0.124	0.129
N	649	69		644	72	

	FSM and IDACI in the top decile					
	Grammar			Independent		
	No	Yes	Yes-No	No	Yes	Yes-No
1+ GCSE with A	0.067	0.062	-0.004	0.096	0.031	-0.064
S.e.	0.024	0.094	0.097	0.025	0.073	0.077
1+ GCSE with A*	0.159	0.125	-0.034	0.064	0.531	0.467
S.e.	0.035	0.128	0.133	0.032	0.142	0.145
5+ GCSE with A-A*	0.180	0.281	0.102	0.128	0.438	0.310
S.e.	0.037	0.144	0.149	0.036	0.148	0.152
N	664	54		661	51	

	No FSM and IDACI in the bottom quartile					
	Grammar			Independent		
	No	Yes	Yes-No	No	Yes	Yes-No
1+ GCSE with A	0.089	0.100	0.011	0.096	0.111	0.015
S.e.	0.027	0.054	0.060	0.026	0.064	0.070
1+ GCSE with A*	0.229	0.141	-0.088	0.160	0.022	-0.137
S.e.	0.038	0.074	0.084	0.035	0.092	0.098
5+ GCSE with A-A*	0.236	0.100	-0.136	0.196	0.089	-0.107
S.e.	0.041	0.076	0.086	0.038	0.103	0.110
N	554	166		626	89	

	No FSM and IDACI in the bottom decile					
	Grammar			Independent		
	No	Yes	Yes-No	No	Yes	Yes-No
1+ GCSE with A	0.087	0.176	0.089	0.093	0.208	0.115
S.e.	0.026	0.079	0.083	0.025	0.085	0.088
1+ GCSE with A*	0.184	0.059	-0.125	0.142	0.083	-0.058
S.e.	0.035	0.111	0.116	0.033	0.138	0.142
5+ GCSE with A-A*	0.207	0.147	-0.060	0.176	0.125	-0.051
S.e.	0.038	0.121	0.127	0.036	0.145	0.150
N	646	67		669	48	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Estimates of CH effect using pscore matching and OLS

	Grammar			Independent		
	ATT	OLS	ATT-OLS % diff.	ATT	OLS	ATT-OLS % diff.
1+ GCSE with A	0.069**	0.046***	-33.2	0.093***	0.077**	-17.5
S.e.	0.023	0.014		0.024	0.015	
1+ GCSE with A*	0.172***	0.194***	12.8	0.127***	0.114**	-10.1
S.e.	0.034	0.021		0.032	0.020	
5+ GCSE with A-A*	0.186**	0.180***	-3.1	0.172***	0.149**	-13.8
S.e.	0.036	0.021		0.035	0.021	
N		6,798			7,560	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Difference in KS1 achievement measures not used to match CH pupils with controls

	Grammar			Independent		
	ATT	Controls		ATT	Controls	
		matched	non-matched		matched	non-matched
KS1 Eng reading task DfE	0.106	-0.091	0.003	-0.003	0.515	0.041
S.e.	0.136			0.076		
KS1 Eng comprehension test DfE	0.100	0.040	-0.015	0.054	0.409	-0.014
S.e.	0.076			0.062		
KS1 Eng writing test DfE	0.024	0.104	-0.014	0.024	0.490	-0.002
S.e.	0.080			0.064		
KS1 Mat using and applying DfE	0.056	0.062	-0.012	0.019	0.453	-0.004
S.e.	0.077			0.067		
KS1 Mat number and algebra DfE	0.038	0.047	-0.014	-0.011	0.460	0.000
S.e.	0.078			0.068		
KS1 shapes and measures DfE	0.068	0.013	-0.010	-0.061	0.481	0.003
S.e.	0.078			0.066		
KS1 Sci enquiry DfE	0.158**	0.019	-0.010	0.101	0.323	0.008
S.e.	0.077			0.069		
KS1 Sci life and living processes DfE	0.106	0.065	-0.016	0.067	0.345	0.001
S.e.	0.078			0.069		
KS1 Sci materials & properties DfE	0.037	0.079	-0.014	0.028	0.329	0.012
S.e.	0.078			0.070		
KS1 Sci physical processes DfE	0.155*	-0.014	-0.012	0.009	0.353	0.007
S.e.	0.081			0.070		
N		988	6,085		1,274	6,844

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Difference in pseudo-outcomes when using only KS2 test achievement and socio-demographics to match CH pupils with controls

	Grammar			Independent		
	ATT	Controls		ATT	Controls	
		matched	non-matched		matched	non-matched
KS1 Eng. reading DfE	0.164**	-0.038	-0.007	0.130**	0.395	0.007
S.e.	0.075			0.059		
KS1 Eng. writing DfE	0.182**	0.003	-0.013	0.094	0.438	-0.004
S.e.	0.079			0.074		
KS1 Mat. DfE	0.033	0.084	-0.015	0.103	0.373	0.005
S.e.	0.077			0.068		
KS1 Eng. read. task DfE	0.050	-0.036	-0.000	0.131	0.381	0.043
S.e.	0.115			0.094		
KS1 Eng. comp. test DfE	0.187**	-0.047	-0.009	0.110	0.353	-0.012
S.e.	0.076			0.062		
KS1 Eng. writing test DfE	0.092	0.037	-0.010	0.088	0.426	0.001
S.e.	0.079			0.066		
KS1 Mat Use and app. DfE	0.106	0.012	-0.010	0.065	0.407	-0.001
S.e.	0.078			0.067		
KS1 Mat num. and alg. DfE	0.027	0.058	-0.013	0.075	0.374	0.004
S.e.	0.078			0.069		
KS1 sha. and mea. DfE	0.031	0.051	-0.013	0.051	0.369	0.009
S.e.	0.078			0.069		
KS1 Sci enq. DfE	0.056	0.120	-0.014	0.127*	0.298	0.009
S.e.	0.076			0.068		
KS1 Sci lif and liv. proc. DfE	0.115	0.056	-0.016	0.106	0.305	0.002
S.e.	0.076			0.069		
KS1 Sci mat. prop. DfE	0.027	0.089	-0.014	0.043	0.315	0.011
S.e.	0.077			0.070		
KS1 Sci phy. proc. DfE	0.090	0.051	-0.016	0.055	0.307	0.009
S.e.	0.077			0.072		
N	.	353.000	6306.000	.	352.000	7272.000

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Difference in KS2 achievement (ATT) when using a rich set of KS1 achievement measures to estimate the pscore

	Grammar			Independent		
	ATT	Controls		ATT	Controls	
		matched	non-matched		matched	non-matched
KS2 Eng TA $\geq$ 4	-0.000	0.997	0.997	0.041***	0.956	0.930
S.e.	0.004			0.011		.
KS2 Mat TA $\geq$ 4	0.000	1.000	1.000	0.023**	0.977	0.930
S.e.	0.000			0.009		.
KS2 Sci TA $\geq$ 4	0.003	0.997	0.997	0.025***	0.975	0.961
S.e.	0.004			0.008		.
KS2 Eng TA > test	-0.039*	0.103	0.110	-0.030	0.094	0.096
S.e.	0.022			0.021		.
KS2 Mat TA > test.	0.019	0.056	0.073	-0.025	0.099	0.082
S.e.	0.019			0.022		.
KS2 Sci TA > test.	-0.006	0.049	0.059	0.019	0.023	0.055
S.e.	0.016			0.014		.
KS2 Eng TA $\neq$ test	-0.062**	0.205	0.209	-0.052*	0.195	0.231
S.e.	0.030			0.029		.
KS2 Mat TA $\neq$ test.	0.039	0.085	0.128	-0.024	0.149	0.164
S.e.	0.024			0.027		.
KS2 Sci TA $\neq$ test.	-0.024	0.125	0.152	-0.045*	0.146	0.206
S.e.	0.025			0.026		.
KS2 Eng score	2.239***	73.592	73.083	3.907***	71.924	66.879
S.e.	0.708			0.840		.
KS2 Mat score	0.353	87.933	86.953	5.738***	82.549	77.985
S.e.	0.657			0.937		.
KS2 Sci score	1.046**	69.469	68.923	2.629***	67.886	65.198
S.e.	0.454			0.551		.
KS2 Eng lev >4	0.081***	0.774	0.750	0.139***	0.717	0.556
S.e.	0.030			0.031		.
KS2 Mat lev >4	-0.003	0.896	0.859	0.173***	0.720	0.624
S.e.	0.023			0.029		.
KS2 Sci lev >4	0.021	0.904	0.870	0.099***	0.827	0.721
S.e.	0.022			0.025		.
N		463	6,314		488	7,108

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Cross-tabulations of achievement in KS1 and KS2 by subject

CH and grammar schools pupils

(a)

		K2 Eng lev.	
		< 5	≥ 5
KS1 Eng lev.	< 3	1,153	1,753
	≥ 3	500	3,392

(b)

		K2 Mat lev.	
		< 5	≥ 5
KS1 Mat lev.	< 3	667	1,780
	≥ 3	269	4,082

(c)

		K2 Sci lev.	
		< 5	≥ 5
KS1 Sci lev.	< 3	605	2,403
	≥ 3	234	3,556

(d)

		K2 1+ lev.	
		< 5	≥ 5
KS1 1+ lev.	< 3	132	1,407
	≥ 3	71	5,188

CH and independent schools pupils

(e)

		K2 Eng lev.	
		< 5	≥ 5
KS1 Eng lev.	< 3	2,668	1,733
	≥ 3	484	2,675

(f)

		K2 Mat lev.	
		< 5	≥ 5
KS1 Mat lev.	< 3	2312	1632
	≥ 3	379	3237

(g)

		K2 Sci lev.	
		< 5	≥ 5
KS1 Sci lev.	< 3	1663	2464
	≥ 3	297	3136

(h)

		K2 1+ lev.	
		< 5	≥ 5
KS1 1+ lev.	< 3	1143	1841
	≥ 3	187	4389

CH and matched pupils in grammar schools

(i)

		K2 Eng lev.	
		< 5	≥ 5
KS1 Eng lev.	< 3	81	161
	≥ 3	37	435

(j)

		K2 Mat lev.	
		< 5	≥ 5
KS1 Mat lev.	< 3	51	169
	≥ 3	22	472

(k)

		K2 Sci lev.	
		< 5	≥ 5
KS1 Sci lev.	< 3	46	227
	≥ 3	9	432

(l)

		K2 1+ lev.	
		< 5	≥ 5
KS1 1+ lev.	< 3	9	121
	≥ 3	2	582

CH and matched pupils in independent schools

(m)

		K2 Eng lev.	
		< 5	≥ 5
KS1 Eng lev.	< 3	68	175
	≥ 3	40	433

(n)

		K2 Mat lev.	
		< 5	≥ 5
KS1 Mat lev.	< 3	59	162
	≥ 3	25	470

(o)

		K2 Sci lev.	
		< 5	≥ 5
KS1 Sci lev.	< 3	43	219
	≥ 3	18	436

(p)

		K2 1+ lev.	
		< 5	≥ 5
KS1 1+ lev.	< 3	9	124
	≥ 3	2	581

Table A.10: Sensitivity analysis of CH effect to all calibrated confounders

	p_{11}	p_{10}	p_{01}	p_{00}	Outcome effect Γ	Selection effect Δ	ATT	S.e.
				Panel A: grammar schools				
				1+ GCSEs with A				
Neutral confounder	0.515	0.517	0.493	0.485	1.031	1.112	0.072***	0.024
				Socio-demographics				
African	0.138	0.172	0.009	0.004	2.886	19.695	0.038	0.030
Caribbean	0.013	0.000	0.002	0.002	1.131	6.205	0.046*	0.027
Bangladeshi	0.000	0.000	0.004	0.003	1.666	.	0.047*	0.026
Pakistani	0.000	0.000	0.010	0.012	1.001	.	0.044	0.027
Indian	0.013	0.034	0.063	0.025	2.795	0.230	0.045	0.028
Other ethnicity	0.242	0.138	0.128	0.096	1.398	2.189	0.044	0.028
Born in 4th trimester	0.188	0.069	0.218	0.265	0.776	0.777	0.047*	0.028
EAL	0.193	0.138	0.115	0.069	1.793	1.910	0.044	0.028
SEN	0.013	0.034	0.023	0.049	0.470	0.542	0.048	0.027
SEN action	0.010	0.034	0.014	0.037	0.375	0.716	0.049*	0.027
SEN action plus	0.003	0.000	0.006	0.010	0.733	0.654	0.049*	0.027
SEN statement	0.000	0.000	0.003	0.002	1.791	.	0.048*	0.026
				KS1				
KS1 Eng. compr. test DfE	0.005	0.107	0.005	0.006	1.200	2.387	0.049*	0.027
KS1 Eng. writ. test DfE	0.315	0.483	0.290	0.183	1.843	1.256	0.046*	0.028
KS1 Eng. spell. test DfE	0.000	0.000	0.000	0.000	.	.	0.074***	0.024
KS1 Mat. task test DfE	0.008	0.069	0.001	0.000	.	12.497	0.045	0.027
KS1 no missing	0.933	1.000	0.962	0.970	0.794	0.597	0.047*	0.027
				KS2 and distance				
KS2 Vol. aided or contr. sch.	0.490	0.552	0.397	0.365	1.143	1.515	0.046*	0.028
KS2 Foundation sch.	0.030	0.000	0.053	0.060	0.892	0.519	0.045	0.028
KS2 Eng. TA > test	0.072	0.042	0.106	0.126	0.831	0.619	0.045	0.027
KS2 Mat. TA > test	0.082	0.000	0.066	0.106	0.595	1.112	0.050***	0.027
KS2 Sci. TA > test	0.045	0.042	0.051	0.103	0.473	0.801	0.050***	0.027
Distance from CH (miles)	0.100	0.138	0.505	0.597	0.689	0.106	0.031	0.028
Closest grammar school (miles)	0.530	0.517	0.496	0.498	0.996	1.149	0.048**	0.027
Closest indep. school (miles)	0.270	0.345	0.494	0.520	0.906	0.387	0.043	0.028
				*1+ GCSEs with A**				
Neutral confounder	0.501	0.553	0.499	0.496	1.014	1.058	0.172***	0.034
				Socio-demographics				
African	0.147	0.105	0.010	0.006	1.838	19.535	0.184***	0.045
Caribbean	0.011	0.013	0.002	0.003	0.828	6.303	0.190***	0.041
Bangladeshi	0.000	0.000	0.005	0.002	2.894	.	0.187***	0.041
Pakistani	0.000	0.000	0.010	0.012	0.899	.	0.185***	0.041
Indian	0.017	0.000	0.068	0.043	1.624	0.221	0.191***	0.042
Other ethnicity	0.238	0.224	0.148	0.091	1.766	2.131	0.182***	0.042
Born in 4th trimester	0.178	0.184	0.213	0.238	0.864	0.768	0.191***	0.042
EAL	0.204	0.118	0.135	0.075	1.951	1.846	0.186***	0.042
SEN	0.011	0.026	0.020	0.035	0.575	0.544	0.186***	0.041
SEN action	0.011	0.013	0.013	0.023	0.559	0.736	0.185***	0.040
SEN action plus	0.000	0.013	0.004	0.009	0.526	0.718	0.184***	0.040
SEN statement	0.000	0.000	0.003	0.003	1.147	.	0.185***	0.040
				KS1				
KS1 Eng. compr. test DfE	0.015	0.000	0.009	0.000	.	2.276	0.185***	0.041
KS1 Eng. writ. test DfE	0.357	0.176	0.366	0.158	3.089	1.198	0.190***	0.042
KS1 Eng. spell. test DfE	0.000	0.000	0.000	0.000	.	.	0.172***	0.034
KS1 Mat. task test DfE	0.015	0.000	0.002	0.000	4.101	12.752	0.188***	0.041
KS1 no missing	0.943	0.908	0.957	0.971	0.681	0.614	0.187***	0.041
				KS2 and distance				
KS2 Vol. aided or contr. sch.	0.513	0.408	0.401	0.383	1.078	1.511	0.193***	0.043
KS2 Foundation sch.	0.025	0.039	0.045	0.065	0.689	0.517	0.189***	0.041
KS2 Eng. TA > test	0.069	0.075	0.084	0.142	0.556	0.636	0.187***	0.042
KS2 Mat. TA > test	0.069	0.119	0.055	0.093	0.569	1.143	0.187***	0.041
KS2 Sci. TA > test	0.033	0.104	0.035	0.088	0.375	0.824	0.189***	0.041
Distance from CH (miles)	0.099	0.118	0.473	0.577	0.659	0.109	0.161***	0.044
Closest grammar school (miles)	0.521	0.566	0.487	0.508	0.919	1.152	0.193***	0.042
Closest indep. school (miles)	0.261	0.342	0.473	0.529	0.797	0.388	0.183***	0.043

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Continued on the next page

Continued from the previous page

	p_{11}	p_{10}	p_{01}	p_{00}	Outcome effect Γ	Selection effect Δ	ATT	S.e.
				Panel A: grammar schools (cont.d)				
				*5+ GCSEs with A-A**				
Neutral confounder	0.508	0.530	0.501	0.491	1.043	1.076	0.188***	0.036
				Socio-demographics				
African	0.161	0.070	0.010	0.006	1.879	19.254	0.175***	0.048
Caribbean	0.006	0.030	0.002	0.002	1.032	6.979	0.192***	0.043
Bangladeshi	0.000	0.000	0.005	0.003	2.344	.	0.188***	0.043
Pakistani	0.000	0.000	0.010	0.011	0.915	.	0.190***	0.044
Indian	0.018	0.000	0.073	0.042	1.824	0.221	0.195***	0.045
Other ethnicity	0.249	0.190	0.147	0.099	1.594	2.129	0.182***	0.045
Born in 4th trimester	0.182	0.170	0.211	0.237	0.868	0.770	0.189***	0.045
EAL	0.216	0.100	0.137	0.080	1.835	1.836	0.183***	0.045
SEN	0.009	0.030	0.021	0.032	0.660	0.542	0.190***	0.043
SEN action	0.009	0.020	0.012	0.022	0.554	0.739	0.189***	0.043
SEN action plus	0.000	0.010	0.005	0.008	0.753	0.674	0.189***	0.043
SEN statement	0.000	0.000	0.004	0.003	1.775	.	0.187***	0.043
				KS1				
KS1 Eng. compr. test DfE	0.013	0.011	0.010	0.000	27.628	2.279	0.187***	0.044
KS1 Eng. writ. test DfE	0.375	0.170	0.381	0.167	3.078	1.193	0.188***	0.045
KS1 Eng. spell. test DfE	0.000	0.000	0.000	0.000	.	.	0.186***	0.036
KS1 Mat. task test DfE	0.016	0.000	0.002	0.000	.	12.064	0.190	0.044
KS1 no missing	0.933	0.950	0.957	0.970	0.696	0.617	0.189***	0.044
				KS2 and distance				
KS2 Vol. aided or contr. sch.	0.489	0.510	0.404	0.381	1.098	1.520	0.189***	0.044
KS2 Foundation sch.	0.024	0.040	0.043	0.065	0.657	0.524	0.190***	0.044
KS2 Eng. TA > test	0.071	0.065	0.086	0.134	0.610	0.642	0.189***	0.044
KS2 Mat TA > test	0.068	0.109	0.056	0.087	0.624	1.152	0.187***	0.043
KS2 Sci TA > test	0.023	0.121	0.035	0.082	0.403	0.857	0.191***	0.044
Distance from CH (miles)	0.109	0.080	0.475	0.562	0.704	0.108	0.157***	0.046
Closest grammar school (miles)	0.517	0.570	0.483	0.510	0.901	1.158	0.188***	0.045
Closest indep. school (miles)	0.255	0.340	0.469	0.527	0.795	0.388	0.178***	0.045
				Panel B: independent schools				
				1+ GCSEs with A				
Neutral confounder	0.480	0.586	0.497	0.513	0.938	0.953	0.093***	0.024
				Socio-demographics				
African	0.138	0.172	0.011	0.022	0.519	12.933	0.118***	0.037
Caribbean	0.013	0.000	0.009	0.033	0.286	0.908	0.105***	0.030
Bangladeshi	0.000	0.000	0.003	0.004	0.954	.	0.107***	0.030
Pakistani	0.000	0.000	0.011	0.008	1.523	.	0.104***	0.031
Indian	0.013	0.034	0.021	0.023	0.974	0.649	0.105***	0.030
Other ethnicity	0.242	0.138	0.159	0.178	0.878	1.615	0.105***	0.032
Born in 4th trimester	0.188	0.069	0.243	0.262	0.904	0.672	0.101***	0.031
EAL	0.193	0.138	0.099	0.111	0.888	2.116	0.105***	0.032
SEN	0.013	0.034	0.059	0.273	0.168	0.135	0.074**	0.030
SEN action	0.010	0.034	0.041	0.134	0.281	0.204	0.090***	0.031
SEN action plus	0.003	0.000	0.014	0.085	0.150	0.163	0.093***	0.031
SEN statement	0.000	0.000	0.004	0.054	0.075	.	0.094***	0.031
				KS1				
KS1 Eng. compr. test DfE	0.005	0.107	0.006	0.006	1.074	2.324	0.105***	0.031
KS1 Eng. writ. test DfE	0.696	0.586	0.527	0.245	3.390	2.428	0.070**	0.030
KS1 Eng. spell. test DfE	0.627	0.760	0.470	0.311	1.959	2.209	0.084***	0.031
KS1 Mat. task test DfE	0.708	0.655	0.548	0.276	3.154	2.395	0.071**	0.030
KS1 no missing	0.933	1.000	0.944	0.930	1.265	0.962	0.104***	0.030
				KS2 and distance				
KS2 Vol. aided or contr. sch.	0.490	0.552	0.401	0.343	1.278	1.529	0.100***	0.032
KS2 Foundation sch.	0.030	0.000	0.041	0.041	1.023	0.692	0.105***	0.030
KS2 Eng. TA > test	0.072	0.042	0.097	0.089	1.102	0.713	0.103***	0.031
KS2 Mat. TA > test	0.082	0.000	0.075	0.105	0.696	0.960	0.104***	0.031
KS2 Sci. TA > test	0.045	0.042	0.049	0.068	0.710	0.871	0.104***	0.030
Distance from CH (miles)	0.438	0.345	0.491	0.540	0.825	0.768	0.101***	0.031
Closest grammar school (miles)	0.388	0.379	0.510	0.492	1.076	0.619	0.105***	0.033
Closest indep. school (miles)	0.470	0.414	0.498	0.514	0.937	0.878	0.103***	0.031

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Continued on the next page

	p_{11}	p_{10}	p_{01}	p_{00}	Outcome effect Γ	Selection effect Δ	ATT	S.e.
				Panel B: independent schools (cont.d)				
				*1+ GCSEs with A**				
Neutral confounder	0.524	0.474	0.500	0.490	1.044	1.090	0.127***	0.032
				Socio-demographics				
African	0.147	0.105	0.008	0.020	0.424	14.016	0.154***	0.046
Caribbean	0.011	0.013	0.008	0.022	0.383	1.003	0.144***	0.038
Bangladeshi	0.000	0.000	0.002	0.004	0.552	.	0.144***	0.038
Pakistani	0.000	0.000	0.010	0.010	1.049	.	0.139***	0.038
Indian	0.017	0.000	0.022	0.021	1.085	0.657	0.141***	0.038
Other ethnicity	0.238	0.224	0.176	0.150	1.218	1.515	0.133***	0.040
Born in 4th trimester	0.178	0.184	0.231	0.266	0.830	0.690	0.133***	0.040
EAL	0.204	0.118	0.107	0.096	1.140	2.003	0.131***	0.041
SEN	0.011	0.026	0.040	0.191	0.184	0.152	0.115**	0.039
SEN action	0.011	0.013	0.028	0.104	0.260	0.225	0.124***	0.039
SEN action plus	0.000	0.013	0.009	0.056	0.171	0.171	0.129***	0.039
SEN statement	0.000	0.000	0.003	0.032	0.086	.	0.135***	0.039
				KS1				
	0.519	0.548	0.458	0.290	2.068	1.590	0.119***	0.039
KS1 Eng. compr. test DfE	0.015	0.000	0.009	0.001	12.526	1.866	0.142***	0.038
KS1 Eng. writ. test DfE	0.727	0.500	0.632	0.264	4.777	2.028	0.094**	0.039
k1 Eng. spell. test DfE	0.667	0.467	0.548	0.290	2.970	1.945	0.106***	0.040
KS1 Mat. task test DfE	0.739	0.536	0.655	0.287	4.746	1.988	0.095***	0.040
KS1 no missing	0.943	0.908	0.943	0.937	1.122	0.978	0.141***	0.039
				KS2 and distance				
KS2 Vol. aided or contr. sch.	0.513	0.408	0.413	0.357	1.270	1.499	0.130***	0.040
KS2 Foundation sch.	0.025	0.039	0.040	0.042	0.967	0.683	0.142***	0.039
KS2 Eng. TA > test	0.069	0.075	0.084	0.107	0.769	0.747	0.140***	0.039
KS2 Mat. TA > test	0.069	0.119	0.071	0.097	0.718	1.007	0.141***	0.039
KS2 Sci. TA > test	0.033	0.104	0.036	0.073	0.483	0.988	0.143***	0.039
Distance from CH (miles)	0.448	0.355	0.468	0.542	0.744	0.798	0.130***	0.040
Closest grammar school (miles)	0.374	0.447	0.492	0.521	0.893	0.627	0.130***	0.040
Closest indep. school (miles)	0.467	0.461	0.472	0.535	0.779	0.914	0.136***	0.040
				*5+ GCSEs with A-A**				
Neutral confounder	0.483	0.450	0.497	0.494	1.013	0.925	0.172***	0.035
				Socio-demographics				
African	0.161	0.070	0.006	0.019	0.329	14.120	0.188***	0.050
Caribbean	0.006	0.030	0.006	0.021	0.312	1.039	0.169***	0.042
Bangladeshi	0.000	0.000	0.003	0.003	1.050	.	0.166***	0.041
Pakistani	0.000	0.000	0.008	0.012	0.697	.	0.164***	0.043
Indian	0.018	0.000	0.026	0.018	1.452	0.622	0.164***	0.041
Other ethnicity	0.249	0.190	0.169	0.160	1.065	1.553	0.154***	0.043
Born in 4th trimester	0.182	0.170	0.226	0.263	0.823	0.698	0.157***	0.043
EAL	0.216	0.100	0.105	0.099	1.072	2.018	0.153***	0.044
SEN	0.009	0.030	0.028	0.173	0.143	0.152	0.133***	0.043
SEN action	0.009	0.020	0.022	0.095	0.218	0.232	0.146***	0.042
SEN action plus	0.000	0.010	0.006	0.050	0.110	0.177	0.152***	0.043
SEN statement	0.000	0.000	0.001	0.028	0.036	.	0.158***	0.043
				KS1				
KS1 Eng. compr. test DfE	0.013	0.011	0.011	0.002	13.169	1.876	0.164***	0.042
KS1 Eng. writ. test DfE	0.743	0.511	0.679	0.297	4.988	1.977	0.112**	0.045
KS1 Eng. spell. test DfE	0.675	0.494	0.585	0.310	3.138	1.876	0.122***	0.044
KS1 Mat. task test DfE	0.775	0.474	0.701	0.320	4.961	1.933	0.117***	0.044
KS1 no missing	0.933	0.950	0.947	0.936	1.215	0.944	0.165***	0.042
				KS2 and distance				
KS2 Vol. aided or contr. sch.	0.489	0.510	0.422	0.361	1.291	1.481	0.151***	0.044
KS2 Foundation sch.	0.024	0.040	0.040	0.042	0.978	0.692	0.165***	0.042
KS2 Eng. TA > test	0.071	0.065	0.078	0.107	0.718	0.769	0.163***	0.042
KS2 Mat. TA > test	0.068	0.109	0.067	0.094	0.692	1.012	0.165***	0.041
KS2 Sci. TA > test	0.023	0.121	0.030	0.070	0.424	1.031	0.167***	0.043
Distance from CH (miles)	0.456	0.350	0.483	0.517	0.873	0.774	0.155***	0.043
Closest grammar school (miles)	0.377	0.420	0.506	0.506	1.003	0.621	0.153***	0.044
Closest indep. school (miles)	0.468	0.460	0.473	0.523	0.818	0.912	0.158***	0.043

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$