HEDG

Health, Econometrics and Data Group

WP 19/08

Human capital consequences of missing out on a grammar school education

Chiara Pastore and Andrew Jones

April 2019

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

Human capital consequences of missing out on a grammar school education.

Chiara Pastore[*‡], Andrew M. Jones[*†]

April 2019

Abstract

What is the value added of grammar schools? This paper disentangles the effect of selection into an academic rather than a vocational track from that of individual background on long-term human capital. Identification relies on a fuzzy regression discontinuity design, using entry test scores for grammar schools, selective secondary schools in England, and estimating discontinuities in school assignment directly from the data. We find that for the marginal admitted student, grammar attendance positively affects educational attainment, likely due to higher-ability peers, while adult labour market outcomes and health are not affected. Observed differences in human capital by school type can largely be traced back to background.

Keywords Selective schooling, Human Capital, Health, Fuzzy regression discontinuity design.
JEL I1, I26, I28, C21.

[*]Department of Economics and Related Studies, University of York, United Kingdom.
[†]Centre for Health Economics, Monash University, Australia.
[‡]Corresponding author: Chiara Pastore. Email: cp798@york.ac.uk. Address: Department of Economics and Related Studies, Heslington, York YO10 5DD, United Kingdom. We thank Nigel Rice and Thomas Cornelissen for helpful suggestions and comments, as well as participants to the 2019 RES Conference at the University of Warwick. Chiara Pastore acknowledges support by the Economic and Social Research Council (grant number ES/J500215/1). Andrew M Jones acknowledges funding from the Leverhulme Trust Major Research Fellowship (MRF-2016-004). The funders, data creators and UK Data Service have no responsibility for the contents of this paper.

1 Introduction

Can being assigned to a school on the basis of ability at age 11 affect long-term outcomes? Does the type of school attended shape differences in life trajectories, or are these differences observed in the raw data already determined by background prior to school? This paper aims at answering these questions by analysing differences in lifetime outcomes by type of school track attended, a dimension of school quality. Proponents of tracking policies, that allocate students to different classes or schools on the basis of ability, maintain that they reward talent regardless of socioeconomic background. Opponents on the other hand raise egalitarian concerns, due to the fact that children from affluent backgrounds are more likely to get better preparation for entry tests and that selection occurs at an early age, thus unfairly penalising children whose ability develops later. In 2018, the UK government announced a 50-million pound fund for an expansion in grammar schools, public and selective high-quality institutions, lifting a ban on the creation of selective school places that had been in vigour since 1998, and thus it is particularly important to look at the long-term effects of selection in education at this time.

Being assigned to a school that selects on the basis of ability might affect long-term outcomes through providing a pool of more able peers, or through a different curriculum, which could facilitate later admission to better higher education programmes. Several elite schools also have direct links with prestigious universities in several countries, and there is evidence that more qualified teachers seek schools with higher ability pupils (Pop-Eleches and Urquiola, 2013). Given the inequality concerns often raised around selective education systems, while measuring the long-term effects of tracking as a dimension of school quality, this paper also indirectly explores whether inequalities are reduced, unaffected or reproduced via a system where pupils are assigned to different schools depending on ability.

In England, admission to grammar schools has traditionally been determined by performance in the 11-plus, an exam taken at age 11 that covered language, numerical and reasoning skills. Until the 1960s the great majority of English children would sit the 11-plus at the end of primary school. Exams were set locally by Local Education Authorities (LEAs), and the entry mark depended on the number of grammar school places in the area. On average pupils scoring in the top 25% of the distribution in their local area were admitted to grammar school. Those who did not reach the entry score were generally assigned to secondary modern schools, with a less academic focus[1]. Thus, students with similar ability scores could be assigned to different types of school for two reasons. Either because they were close to the cutoff entry score for their area, or because they were

[1]In 1965 a third type, comprehensive schools, was introduced, catering for students of all abilities. This paper focuses on areas that were still largely selective when the data was collected, and where almost the totality of pupils attending public secondary schools was assigned to either grammar or secondary modern.

from different areas. To the extent that scoring slightly above or below the entry cutoff is random, this paper can isolate the long-term effect of going to grammar school, compared to just missing out on admission, on human capital for individuals in the National Child Development Study (NCDS). This is a cohort of English individuals born in March 1958 who started secondary school in 1969, and whose lives have been followed for 60 years to date.

We use a regression discontinuity design (RDD) based on the fact that grammar admission was determined by whether the 11-plus score exceeded a local entry score threshold. As a proxy for the 11-plus score, we use age 11 cognitive tests collected in the NCDS. These closely mirror the three components of the 11-plus and are therefore reliable predictors of grammar school entry. Since, to the best of our knowledge, information on the entry score cutoffs for all English LEAs in the 1960s is not available, we proxy cutoffs with LEA-specific thresholds estimated directly from the data. Following the structural breaks literature (Bai, 1997), in the same spirit as Card, Mas, et al. (2008), we select the threshold value that maximises the fit of a simple model of school assignment. We account for possible sampling error in threshold estimation by using bootstrapped standard errors. This approach identifies the medium-term effects on educational achievement and labour market outcomes, as well as long-term effects on health and risk of developing disease up to age 50. The first stage is strongly predictive of grammar attendance, indicating confidence in the empirical strategy. In the second stage, for the marginal student, we find a significant and positive effect of grammar attendance on educational attainment. No effect is found on adult labour market outcomes, health and biomarkers for risk of developing chronic diseases. This result holds both for a standard fuzzy RDD approach and for bias-corrected procedures proposed by Calonico, Cattaneo, Farrell, and Titiunik (2017) and Calonico, Cattaneo, and Titiunik (2014a). We acknowledge that compared to the 1960s there are several types of school granting access to A-levels and university nowadays, meaning the grammar advantage in educational outcomes may not be as important to this day. We also find that peer quality appears to be a significant mechanism to achieve better educational outcomes, although a large portion of human capital differences by school type is due to background selection, and more specifically to pre-schooling ability and parental background.

RDD based on assignment test scores has been a popular approach to estimate the causal effect of higher-quality schools on a variety of outcomes in several countries (Abdulkadiroglu et al., 2014; Del Bono and Clark, 2016; Dobbie and Fryer, 2014; Dustan, 2010; Dustmann et al., 2016; Guyon et al., 2012; Kirabo Jackson, 2010; Pop-Eleches and Urquiola, 2013; see Table 1 for more details). Most studies find a positive effect of higher-quality schools on measures of educational attainment. Some however, find no effect (Abdulkadiroglu et al., 2014; Dustmann et al., 2016) or even a negative effect, as in the case of a Swedish reform that made vocational education more academic, which was found to increase drop out rates (Hall, 2012). A highly relevant study by Clark (2010), who implements a regression discontinuity design to investigate the impact of grammar schools on educational outcomes for Yorkshire, a region in England, finds no effect on educational achievement, but a positive effect on university enrolment. Del Bono and Clark (2016) and Dustmann et al. (2016) find no significant effect of higher academic ability schools on labour market outcomes such as earnings, probability of employment, occupational choice, with the exception of a small positive effect on women's earnings in Del Bono and Clark (2016). These studies stand out in the literature because of their focus on long-term outcomes other than educational attainment. Dustmann et al. (2016) study the German secondary school system, exploiting month of birth as a source of quasi-random variation in track assignment at age 10. They relate the absence of long-term effects of attending the low, middle or high track to flexibilities in the system, which allow for reallocation of students to a more suitable track at later grades. Del Bono and Clark (2016) is the study most similar to ours, since they investigate long-term effects of elite schools in Aberdeen, Scotland, for children born in the 1950s. In addition to educational and labour market outcomes, they study effects on fertility. Their strategy exploits information on admission exam scores (equivalent to the 11-plus), and entry score cutoffs for the city of Aberdeen, for an RDD-type strategy.

Given the unavailability of such precise information at Local Education Authority level for all of England, literature on the effects of grammar schools has been unable to exploit RDD methodologies to answer this question in the English context, with the exception of the paper by Clark (2010), which focuses on one region in England and on education outcomes only. The most popular approach has been instrumental variables, used to compare selective and comprehensive systems, a practice that has been challenged in the literature. Manning and Pischke (2006) argue that instrumental variables are unable to remove the selection problem arising from fundamental differences between areas adopting different systems. This paper thus contributes to the literature on the long-term effects of grammar schools in England in several ways. We are the first to exploit an RDD strategy to answer this question for all of England, using data from a nationally representative study. Second, we implement RDD in a context with limited information, constructing the assignment variables and score cutoffs from large survey data, in the absence of administrative records. We use several robustness checks that are standard in RDD, increasing confidence in the robustness of the estimation strategy and further confirming previous results obtained with other methodologies. Moreover the NCDS dataset includes high-quality information on individuals' pre-treatment characteristics, allowing us to analyse background as a competing explanatory variable as well as to check similarity between treated and control group near the threshold. Third, we are able to investigate a broad range of long-term outcomes rather than only educational ones, including long-term health conditions and disease risk. Also, since schools are surveyed as part of the study, we are able to investigate how specific school features may serve as channels for long-term outcomes. Finally, the issue of selective schools is highly topical, having repeatedly cropped up in the political arena in England since the 1960s, and continuing to be the subject of educational policy debates to this day in several other countries (Adams, 2018; Jeffreys, 2018).

The paper is structured as follows. We describe the data in Section 2. Section 3 discusses the RDD set up, followed by the necessary conditions for identification and a discussion of the empirical approach to estimation of a causal effect. Section 4 presents the results, along with robustness checks and an exploratory pathway analysis and it is followed by a discussion and concluding remarks in Section 5.

2 Data

The NCDS is a longitudinal study of individuals born in the United Kingdom in a single week in March 1958. 98% of all individuals born in England, Scotland and Wales during that week were part of the birth survey, making it a nationally representative sample. Following the first sweep at birth, which contained over 17,000 individuals, surveys were also undertaken at ages 7, 11, 16, 23, 33, 42, 45, 50, 55. At the latest sweep the survey still retained over 9,000 individuals (Brown et al., 2016)[2].

Information on grammar school attendance is retrieved from the age 16 wave. The sample consists of individuals who went to grammar or secondary modern schools between the ages of 11 and 16, attended by 10% and 20.6% of the NCDS sample respectively. The main variable needed for the identification strategy, age 11 cognitive ability scores, is obtained via principal component analysis (PCA) for the maths, reading and general ability test modules included in the NCDS, following previous literature (Cawley et al., 1997; Galindo-Rueda and Vignoles, 2005; Jones, Rice, and Rosa Dias, 2011). These components and their relative weights in the index mirror the 11-plus components and weights, offering support to the assumption that NCDS ability scores are good proxies of 11-plus scores. Information on LEA of school, essential for recreating the actual peer group of test-takers from NCDS data, was obtained via special licence access. The sample for threshold estimation thus consists of grammar and secondary modern pupils for whom we also have age 11 cognitive test scores, which yields 3,448 individuals. Due to the inclusion of covariates, and to missing items from surveys at different ages, samples for outcome regressions are always smaller, ranging roughly between 1,450 and 2,800, depending on the outcome[3].

[2]Attrition over time can be a problem for treatment effect estimation when it is correlated to specific individual characteristics. Previous literature has argued that this is not necessarily a concern for the NCDS in this setting, especially given that we control for concerning characteristics in the model specification (Case et al., 2005; Dearden et al., 2002; Jones, Rice, and Rosa Dias, 2011).

[3]The reason why we do not use the same sample for threshold estimation and estimation of treatment

Our broad range of outcomes allows us to build a rounded picture of the consequences of school quality for the individual, covering the education, labour market and health domains. Education outcomes are dummies for any A-levels achieved and having a university degree, asked at age 23. Labour market outcomes are measured at age 33 and are all retrieved from survey questions. They include dummies for unemployment and receiving state benefits (excluding child benefits), as well as gross hourly wage, which is calculated from weekly hours worked and weekly wages and then log-transformed for regression analysis. Health outcomes include binary variables for self-assessed health (SAH, equal to 1 if excellent or very good, 0 otherwise), and low malaise score (equal to 1 if the person presents at most 2 out of 9 ill mental health items, 0 otherwise), both evaluated at age 50. Body mass index (BMI) is calculated as weight in kg divided by squared height in meters, and it is measured at age 45. Cholesterol ratio and triglycerides are retrieved from blood samples also taken at age 45. The three measures are all increasing in the risk of cardiovascular disease and health complications (Benzeval et al., 2014).

The main advantage of cohort studies such as NCDS, compared to administrative data, is that we can control for specific individual characteristics measured very early on in childhood. We account for childhood non-cognitive skills in all regressions, given their demonstrated importance for long-term outcomes (Kautz et al., 2014). They are measured at age 11, prior to starting secondary school, by asking primary school teachers questions on the twelve behavioural dimensions that are part of the Bristol Social Adjustment Guide (BSAG). We then converted the total BSAG score to a variable bound between 0 and 1 for convenience. The vector of covariates further includes sex; mother's interest in child education and father's socioeconomics status on 5-point scales; whether the mother was smoking at the fourth month of pregnancy; a childhood morbidity index for the cohort member. The morbidity index was constructed following previous literature, by summarising information on twelve categories of childhood conditions up to age 7 into a variable bound between 0 (no morbidity) and 1 (highest morbidity).

3 Methods

3.1 Framework for identification

We exploit a discontinuity in the treatment assignment function to estimate long-term effects of attending grammar school in childhood, compared to going to a secondary modern, for a range of adult education, labour market and health outcomes. In this setting the discontinuity is due to the fact that grammar school admission was largely determined by a LEA-specific pass mark, so the assignment variable is the entry exam score (or 11-

effect is that a reliable threshold figure needs to capture as much of the actual test-taking population as possible.

plus score), here proxied by the age 11 cognitive ability score index obtained via PCA as described in Section 2. Identification of treatment effects in the RDD is based on the assumption that pupils scoring near the pass mark, where the discontinuity is observed, have very similar baseline characteristics, so that near this threshold treatment assignment is as good as random, and differences in long-term outcomes are caused exclusively by treatment (Lee and Lemieux, 2010). Estimating the extent to which treatment alone causes these differences yields a local average treatment effect (LATE) for the group of compliers. These are the individuals in proximity of the threshold, who are assigned to treatment by virtue of scoring above the cutoff.

The first stage of the empirical strategy models school assignment. We denote the treatment variable as $G \in \{0, 1\}$, where $G = 1$ indicates grammar attendance. Following Lee and Lemieux (2010), treatment assignment, which is assumed to change discontinuously at a LEA-specific cutoff level c_{LEA} of the assignment variable A_i, ability test score, is modelled as:

$$G_i = \gamma + \theta 1[A_i \geq c_{LEA}] + h(A_i - c_{LEA}) + \upsilon_i, \tag{1}$$

where $1[A_i \geq c]$ is an indicator equal to one if the value of the assignment variable is equal to or greater than the threshold, $h(.)$ is a generic function of individual's distance from the pass mark and υ a random error term. Note that we expect this function to be non-deterministic in the sense that children with the same score may be assigned to different schools. There are at least two reasons why we see this in this context. First, because of imperfect compliance by the school, due to the fact that cognitive test scores by themselves did not grant access to grammar school. In particular, admission also depended on area-specific characteristics, such as school capacity constraints, and individual-specific characteristics, namely distance from the school and having other siblings already at the school among others. Probability of accessing grammar is then a function of where the pupil ranks in term of cognitive ability, relative to other children of their age, and other variables. Second, additional fuzziness may be caused by limitations in the data, as noted by Card and Giuliano (2016) when analysing US school data. Recall that we are using a proxy of actual 11-plus scores, and children may have performed differently in NCDS tests than in the 11-plus. Additionally, information on LEA-specific pass marks applied in 1969 is not available to the best of our knowledge. Although we provide evidence in support of the threshold selection procedure performed, we acknowledge that it may further increase fuzziness in the first stage (details in Section 3.2.1). In consideration of these characteristics of the treatment assignment function, we rely on a fuzzy RDD, where the probability of treatment assignment does not jump sharply from 0 to 1 at the cutoff value of the ability test score A_i, but by a smaller amount, for the reasons listed.

The second-stage equation characterising the fuzzy RDD can then be expressed as:

$$Y_i = \alpha + \beta G_i + f(A_i - c_{LEA}) + \epsilon_i. \tag{2}$$

where Y_i are human capital outcomes, β the treatment effect of interest, $f(.)$ a function of distance from the threshold and ϵ a random error term. The vector of individual-level covariates X_i has been suppressed for ease of notation, but these are assumed to enter both Equations 1 and 2. Following previous literature, we propose an RDD estimator analogous to a Wald estimator in 2SLS procedures (Hahn et al., 2001; Lee and Lemieux, 2010), so that average treatment effect is identified by the change in the outcome variable produced by a change in the assignment variable (i.e. the reduced form), divided by the change in the first stage, expressing treatment as a function of the assignment variable.

Interpretation of the Wald estimator as an average treatment effect is conditional upon the following assumptions. The first necessary condition is that the assignment variable cannot be precisely manipulated by the individuals in the sample. In this setting we use a proxy for 11-plus test scores, which cannot be precisely manipulated by pupils. Precise manipulation of the assignment variable seems unlikely, since individuals have no incentive to change their NCDS ability score depending on the local area grammar school pass mark. In standard RDD a condition in support of the no precise manipulation assumption is that the density of the assignment variable should be reasonably smooth around the threshold, routinely tested via McCrary tests for the assignment variable (McCrary, 2008). The test looks for discontinuities in the density function of the assignment variable, the absence of which supports the smooth density assumption. However in the present case the test may be inadequate, since the LEA-specific threshold is estimated based on a goodness of fit criterion, and we would expect this to be reflected in the density function of the constructed assignment variable. Instead, we check whether our main results are sensitive to excluding a portion of selected observations around the threshold, an approach known as donut-RDD (Barreca et al., 2016).

The second necessary condition is that other pre-treatment covariates are smooth functions of the assignment variable, to rule out that the treatment effect estimate is confounded by discontinuities in other variables. Following Card and Giuliano (2016), we test this by constructing a covariate index from regressing each outcome on all pre-treatment covariates of interest, and plot fitted values against the assignment variable, grouping them into fixed-size bins. The aim is to look for jumps in the average index, which would invalidate unbiasedness of the treatment effect estimate. The index graphs in Figure 1, showing fitted values for all outcomes, do not contradict the covariate smoothness assumption, meaning the discontinuity in treatment assignment is only caused by the assignment variable.

Under monotonicity of the instrument (i.e. $\theta > 0$ for all i, or $\theta < 0$ for all i)

and the other verified assumptions (no precise manipulation, smoothness of the density of the assignment variable and smoothness of the assignment variable as a function of covariates), β can formally be interpreted as a local average treatment effect (LATE) for individuals in the proximity of the threshold (Lee and Lemieux, 2010). That is, $\beta = E\{Y_i(1) - Y_i(0)|A_i = c\}$, where $Y_i(1)$ and $Y_i(0)$ denote treated and untreated outcomes respectively. As specified above, the LATE is calculated for compliers, who are those individuals who attend grammar rather than secondary modern because their score allows them to be just above the cutoff for their LEA. Estimating the long-term effect of grammar attendance for this group is interesting because it allows us to understand the effect of selection for students whose ability is almost good enough for grammar schools, and who are allocated to a lower-quality school because of school capacity constraints (i.e. last available places were filled up by someone scoring just above them). At the same time, under the stated assumptions the estimation strategy allows us to isolate the effect of grammar from pre-schooling ability and other pre-treatment confounders. However, we note that it may underestimate the total effect of grammar: it may be that the benefit of grammar schools is concentrated among students at the very top of the ability distribution, but here the estimates are not apt to capture those effects.

3.2 Implementation

3.2.1 Threshold selection

The threshold for each local education authority (LEA) is estimated directly from the data. We do not have information on actual grammar admission marks by LEA, but we can infer them from the ability test scores of pupils attending grammar and secondary modern schools within each LEA in the NCDS. Asymptotic theory for estimated discontinuity points comes from the financial structural breaks literature (Bai, 1997; Bai et al., 1998) and more recently it has also been developed in broader settings (Porter and Yu, 2015). Reassuringly, the key finding from this literature for the present application is that treatment effect in the RDD can be efficiently estimated in presence of discontinuity points estimated from the data (Porter and Yu, 2015). Thus, for each LEA we first run probit models for grammar attendance with a single regressor $1[A_i \geq c]$, indicating whether ability A_i is above a threshold c, for a pre-specified range of possible thresholds $c \in [-0.2, 1.5]$. A grid search for the highest log-likelihood achieved by these models for each LEA then yields the chosen LEA-specific threshold c_{LEA}.

The approach is close in spirit to Card, Mas, et al. (2008), who look for the presence of tipping points (i.e. discontinuities) in a function expressing changes in the share of white population over time at neighbourhood level in the United States. Their hypothesis is that discontinuities in these changes are located at specific values of the base-year minority population share. The study analyses a much larger sample than ours, and

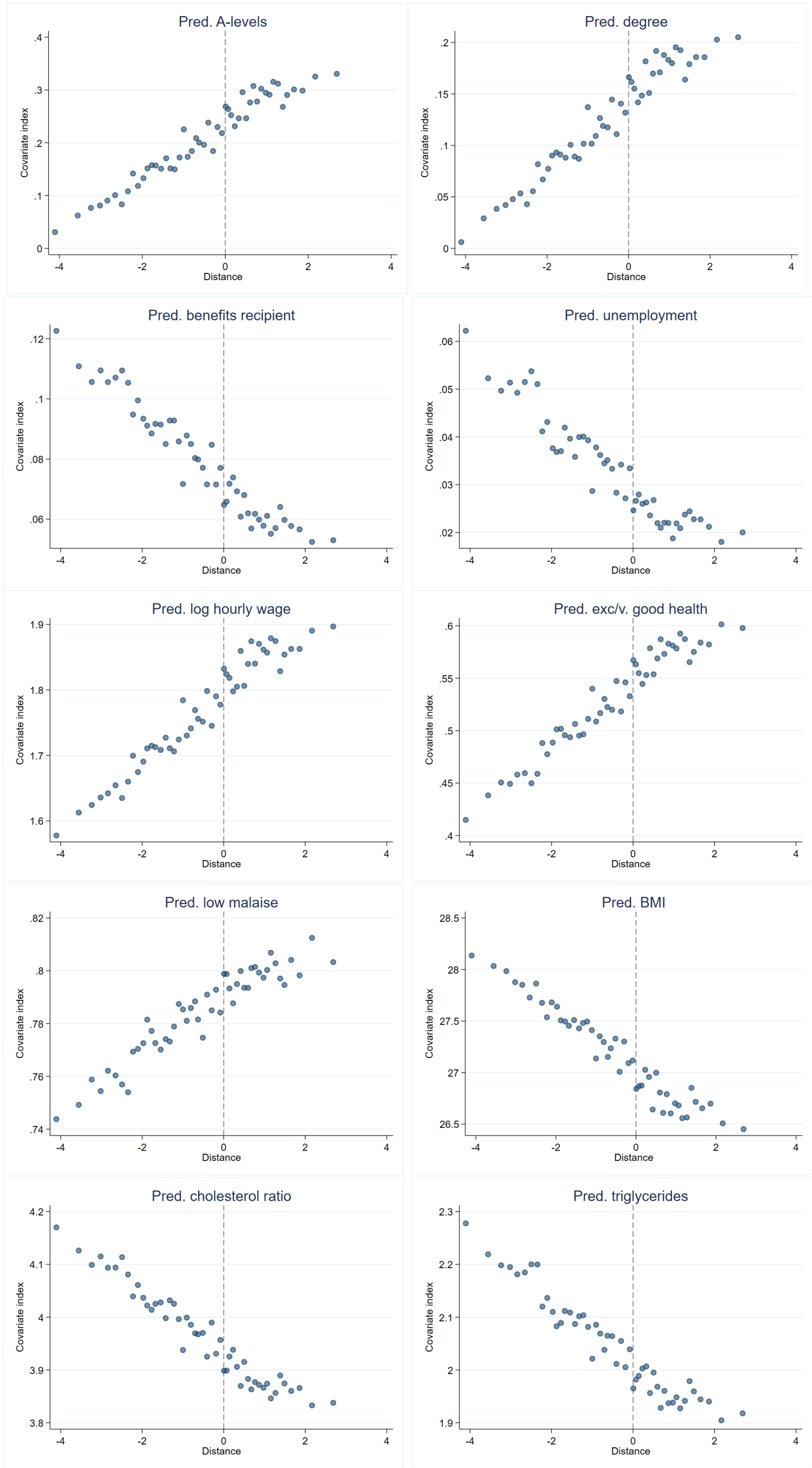

Figure 1: The covariate indices for predicted outcomes are smooth functions of the assignment variable around the threshold.

can therefore estimate the threshold on a subsample of the total available sample, and then use the estimated tipping point on the remaining observations, reducing the risk of sampling bias[4]. Given the smaller sample size in the present case, we do not split the sample, but we bootstrap the threshold search procedure, simulating the estimation sample and subsequent threshold search 500 times, and then use bootstrapped standard errors for statistical inference for the treatment effect estimators. Notice that compared to Card, Mas, et al. (2008), we have the advantage that an LEA-specific discontinuity was actually present in the school assignment function (since there was a pass mark for grammar school entry), although its exact location is unobservable to us and our estimate is based on a relatively small sample size.

3.2.2 Choice of bandwidth, kernel and polynomial

As previously stressed, the validity of the RDD to estimate a LATE for compliers relies on similarity of individuals in the proximity of the threshold, a situation that mimics random treatment assignment. A way to ensure similarity is to accurately choose the neighbourhood around the cutoff in order to select the observations for the estimation of β. Following the RDD literature, we refer to this neighbourhood as the bandwidth h (Lee and Lemieux, 2010). In practice, neighbourhood selection is affected by the choice of bandwidth, as well as kernel function and polynomial order for the local first and second stage regressions.

The smaller the bandwidth, determining the window h below and above the threshold where observations can be drawn from for estimation, the higher the number of observations excluded, and the higher the probability that the similarity assumption holds for individuals whose assignment variable lies within $[c-h, c+h]$. Bandwidth selection then incurs a trade-off between precision and bias, since larger windows around the cutoff will yield estimates with lower variance but potentially higher bias (Lee and Lemieux, 2010)[5]. A popular approach is to choose the bandwidth that minimises an approximation of the mean squared error (MSE) of the local linear estimator of β, $MSE = (\hat{\beta} - \beta)^2$ (Calonico, Cattaneo, and Titiunik, 2014a; Imbens and Kalyanaraman, 2012).

While MSE-optimal bandwidth is generally recommended for its point estimator performance, a recent body of literature has shown it is inadequate for inference procedures (Calonico, Cattaneo, Farrell, and Titiunik, 2017; Calonico, Cattaneo, and Titiunik,

[4] Alongside the 'structural break' approach we use, Card, Mas, et al. (2008) also implement a 'fixed point' method, based on finding the unit root of the polynomial expressing the first stage. Since we are not working with changes, where zero can be a saddle point in the polynomial function, but with the probability of attending grammar, this approach is not a viable option here.

[5] Bias arises because the further from the threshold, the larger the differences between individuals, not only due to treatment but also due to other confounders. Hahn et al. (2001) note that bias of the RDD estimator can be substantial in finite samples, because of the bad behaviour of kernel regression estimators at boundary points, which is a common problem in nonparametric estimation, but seems especially severe for RDD-type settings, since all estimation points are at a boundary.

2014a; Calonico, Cattaneo, and Farrell, 2018; Calonico, Cattaneo, Farrell, and Titiunik, 2018; Cattaneo and Vazquez-Bare, 2016). The argument in a nutshell is that in using the estimated MSE-optimal bandwidth for neighbourhood selection (i.e. $[c-h_{MSE}, c+h_{MSE}]$) we are introducing a misspecification bias, but this bias makes inference based on observations within the neighbourhood and the resulting point estimator invalid (Cattaneo and Vazquez-Bare, 2016). Since the MSE-optimal bandwidth is usually too large for inference, one possible approach would be to simply shrink it, known as undersmoothing (Cattaneo and Vazquez-Bare, 2016; McCrary, 2008). Calonico, Cattaneo, and Titiunik (2014b) instead propose a 'robust bias correction' procedure for bandwidth selection, which estimates bias and then adjusts both the point and variance estimators. The bias correction allows construction of robust confidence intervals, accounting for possible bias and estimation error in both the treatment effect and bias estimates[6]. In the two main specifications we compare local regressions for outcomes with MSE-optimal bandwidth selection approach without and with bias correction, in order to investigate sensitivity of the analysis to bias correction[7]. For a discussion of other possible criteria for bandwidth selection, including cross-validation (Lee and Lemieux, 2010), Fan and Gijbels' (1996) bandwidth selector and local randomization neighbourhood selection methods, see Cattaneo and Vazquez-Bare (2016).

The RDD requires a kernel function to be chosen. The kernel function assigns weights to observations depending on their distance from the threshold, in order to provide optimal treatment effect estimates. While a triangular kernel, assigning highest weight to observations near the threshold, is intuitively appealing, both Lee and Lemieux (2010) and Card and Giuliano (2016) find no important efficiency losses from using uniform kernels. Since uniformity also makes for a simpler computation and interpretation of results, we rely on bandwidth selection to select similar observations, and use a uniform kernel for the main strategy, thus giving equal weight to all observations.

A second choice is order of polynomial to be used, given that introducing higher order terms for distance from the cutoff often improves fit of the first stage regression. However, since recent literature has shown that regression discontinuity analysis based on high-order polynomials may be misleading (Gelman and Imbens, 2018), we use a 1st-order polynomial, and only introduce an interaction between threshold and distance from threshold in the main specification. This accounts for the intuition that not only

[6]The bias estimator and the theory behind it can be found in Calonico, Cattaneo, and Titiunik (2014b), with a more recent take in Calonico, Cattaneo, and Farrell (2018). An alternative approach to MSE-optimal bandwidth, introduced by Calonico, Cattaneo, and Farrell (2018), aims at minimizing the coverage error (CE), which stems from selecting individuals whose characteristics are dissimilar, thus biasing treatment effect estimates. Cattaneo and Vazquez-Bare (2016) and Calonico, Cattaneo, and Farrell (2018) suggest using bandwidth h_{MSE} for point estimates, while h_{CE} for more robust confidence intervals, since generally $h_{MSE} > h_{CE}$, and estimation based on the latter would lead to too much variability in the estimate.

[7]The bias-corrected estimates are obtained with the 'rdrobust' package for Stata (Calonico, Cattaneo, Farrell, and Titiunik, 2017; Calonico, Cattaneo, and Titiunik, 2014b).

the intercept but also the slope of the average of the dependent variable as a function of the assignment variable may be different after the threshold[8].

The empirical counterparts to equations 1 and 2 are then as follows:

$$G_i = \gamma + \theta_0 T_i + \theta_1 D_i + \theta_2 D_i \times T_i + X_i'\eta + \upsilon_i \tag{3}$$

$$Y_i = \alpha + \beta_0 G_i + \beta_1 D_i + \beta_2 D_i \times T_i + X_i'\xi + \epsilon_i, \tag{4}$$

where $T_i = 1[A_i \geq c_{LEA}]$ is an indicator for above the threshold, $D_i = (A_i - c_{LEA})$ indicates the distance between the individual's ability test score in the NCDS and the LEA-specific threshold, $D_i \times T_i$ is the interaction between these two and X_i is a vector of individual characteristics. We estimate equations 3 and 4 as the first and second stage of a two-stage least squares regression, with G_i as the endogenous treatment variable, and only on the sample selected by using the MSE-optimal bandwidth. We present results without bias correction first, and then with bias correction as a robustness check, following the procedures proposed by Calonico, Cattaneo, and Titiunik (2014b).

We include a battery of robustness checks of our approach. We first implement RDD regressions for a placebo cutoff, in order to check that first stage is not predictive of grammar attendance at other points of the distribution of the assignment variable. Secondly, we re-estimate the model with and without bias correction while excluding a set of observations around the threshold, a procedure known as donut exclusion or donut-RDD, to exclude that irregularities in the density function around the threshold might be driving our results. Thirdly, we show that our approach successfully isolates the effect of grammar school attendance from that of background by estimating the effect of the discontinuity in grammar attendance on placebo outcomes prior to secondary school, namely maths scores and BMI at age 7. This check is in the same spirit as an influential test conducted by Manning and Pischke (2006), showing that background confounders were not appropriately accounted for in value-added approaches, since outcomes prior to secondary school appeared to be affected by secondary school attendance[9].

3.3 Mechanism analysis

Finally, we provide a brief discussion of possible channels for long-lasting effects of grammar attendance on human capital. The idea is to pick out specific attributes of grammar schools that may lead to more favourable outcomes for grammar pupils. Data on school attributes comes from the school questionnaire that was part of the age 16 NCDS wave. Socioeconomic status of other school children is proxied by the proportion of fathers in

[8]For formal details on the choice of bandwidth, kernel and polynomial and the squared error minimization procedure with the inclusion of covariates see Calonico, Cattaneo, Farrell, and Titiunik (2017) and Calonico, Cattaneo, Farrell, and Titiunik (2018)

[9]We are grateful to Sandra McNally for suggesting this check.

non-manual occupations and by the share of children that stayed at school beyond minimum school-leaving age before 1972, year of the reform raising compulsory schooling from 15 to 16 years of age. Peer ability is measured by looking at percentages of pupils taking either only GCEs (General Certificate of Education or O-levels, of higher academic value) or only CSEs (Certificate of Secondary Education, of lower academic value, not requiring completion of a full standard qualification)[10]. Finally, we use indicators for whether the school lacks a library, science laboratories or sports facilities as proxies of school resources.

This exercise is mainly exploratory in nature, in the sense that we do not provide a breakdown of the components of the full treatment effect. Rather, we provide a first evaluation of the extent to which each mechanism might explain the long-term effect of a grammar school education. We dichotomise the variables for peers' socioeconomic status, school children ability and school resources, so we can reproduce the main analysis, substituting treatment in the first stage by each of the channels of interest in turn. We adopt the following empirical specification:

$$M_i = \pi_0 + \pi_1 T_i + \pi_2 D_i + \pi_3 D_i \times T_i + X_i' \kappa + \upsilon_i^M \tag{5}$$

$$Y_i = \mu_0 + \mu_1 M_i + \mu_2 D_i + \mu_3 D_i \times T_i + X_i' \nu + \epsilon_i^M, \tag{6}$$

Since we are taking the cutoff for grammar attendance as instrument in the first stage, the estimated parameter μ_1 gives an indication of how important each channel might be for determining outcomes for the population of compliers described in Section 3.1.

4 Results

4.1 Summary statistics

Summary statistics for baseline characteristics and outcomes by type of school for the whole sample are shown in Table 2. Grammar pupils display higher average skills, a higher proportion of female pupils and a more advantaged parental background than secondary modern pupils, while no difference is shown in the childhood morbidity of the two groups. Grammar pupils also display better outcomes in adulthood across all domains. They have a much higher chance of getting A-levels (50% vs 3%) and a university degree (31% vs 2%); they are half as likely to be unemployed or have received state benefits at 33 (excluding child benefits), and they have a higher hourly wage. The long-term health of grammar pupils is also better. At age 50 they are more likely to display high levels of

[10]CSEs were introduced in 1965 in order to provide a certification for students who were leaving school at 16 without a formal secondary school qualification (see webpages at https://qualifications.pearson.com).

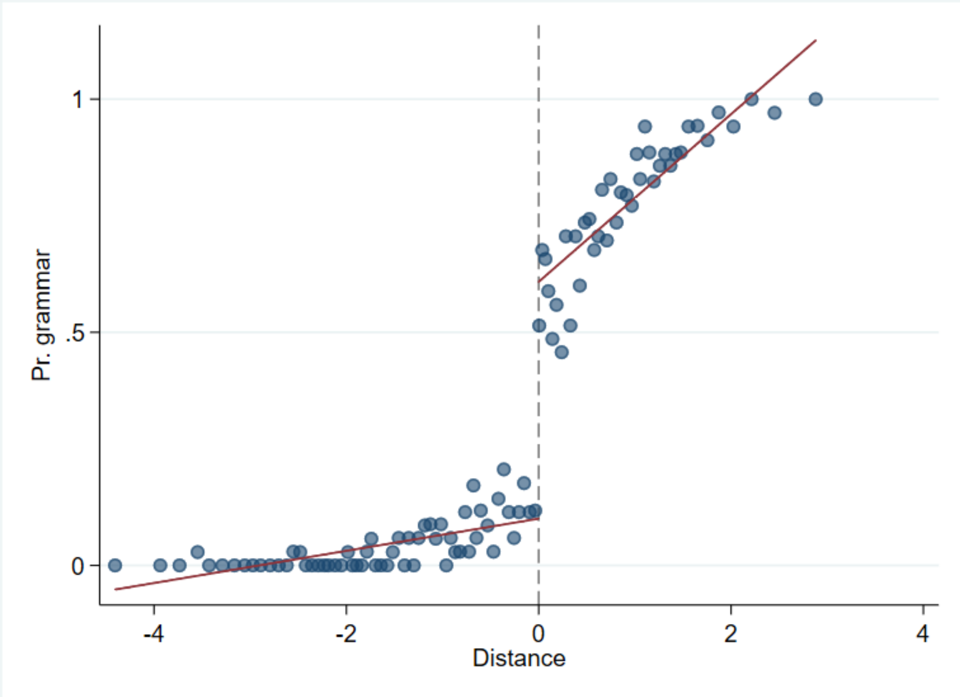

Figure 2: Scatter graph for probability of treatment as a function of the distance between the assignment variable A_i and the LEA-specific threshold. Observations are grouped in 50 bins, yielding a raw figure for average probability of treatment within the group. Average bin size is 69 (N=3448).

self-assessed health (SAH) and low levels of malaise. Their risk of cardiovascular disease and comorbidities at age 45 is also lower, as shown by the lower average levels of BMI, cholesterol ratio and triglycerides.

As a preliminary analysis we conduct OLS regressions for each outcome with grammar as a single regressor and then adding controls, using the whole sample of pupils. Table 3 shows that the coefficient for grammar is highly significant for all outcomes, except for the malaise dummy. However, in line with expectations, some of this association is accounted for by adding other variables, mainly cognitive skills, but also sex, mother's interest in child education and father's socioeconomic status. The RDD treatment effect estimate we present below aims at isolating the effect of grammar for the group of individuals who are very close to the threshold, netting out the effect of cognitive skills as well as other factors, since in proximity of the threshold these should not vary significantly.

4.2 Discontinuity in grammar assignment

Figure 2 plots grammar attendance in pre-specified bins for different levels of the distance between the assignment variables and the LEA-specific cutoff. The figure shows a jump in the probability of treatment when the assignment variable is above the threshold, that is, when the distance between the ability test score A_i and the LEA-specific cutoff c_{LEA} is zero. After this threshold, the average probability of treatment increases by approximately 0.4-0.5, and it keeps rising up to 1 for the most able individuals. This holds regardless of whether the graph is plotted for the whole sample, or separately by gender. Something similar is observed by Dong (2018) in respect to the data used by Del Bono and Clark (2016), who also estimate the impact of elite schooling, but in Scotland and on different outcomes. Dong (2018) adopts a regression kink design (RKD) in re-analysing the Scottish data, but Figure 2 shows a clear discontinuity in treatment assignment in our application.

Full results for the first stage estimated using the samples for each of the 10 outcomes in turn, and including all covariates, are shown in the Appendix (Table A3). Effects for the instrument of interest, the dummy for ability score above the LEA-specific threshold, $1[A \geq c_{LEA}]$, are sizeable, all pointing to an approximate 0.45 increase in the probability of attending grammar at the discontinuity, as already shown by graphical evidence for the whole sample. The distance variable is only significant when interacted with the threshold indicator, and only for the education and labour market outcomes and for the malaise indicator. This makes sense, since below the threshold the probability of attending grammar is close to zero, regardless of the distance. The probability of attending grammar is also positively and significantly associated with non-cognitive skills, mother's interest and father's SES for most of the samples considered.

4.3 Effect of grammar for the marginal student

Figure 3 shows scatter graphs for the outcomes for the whole sample, so that each dot represents average outcome for groups of individuals at similar levels of the distance variable. All outcomes display a relationship with the assignment variable in the expected direction. However no outcome shows a sharp jump at the threshold value of zero (indicated by the dashed line). The probability of achieving A-levels and a degree display a fairly similar pattern, increasing continuously after the threshold. Most other outcomes show an average improvement of some sort, more evident for cholesterol ratio, triglycerides and the indicator of excellent or very good SAH.

Tables 4 and 5 show RDD estimates for education, labour market and health outcomes obtained via estimation of Equations 3 and 4, estimated for the sample within MSE-optimal bandwidths, without bias correction. Standard errors are bootstrapped to account for potential noise in the estimation of the threshold. All models include all covariates, as described in Section 2. The difference in total available sample across outcomes depends on missing information for certain outcomes, and on the fact that MSE-optimal bandwidths vary for each outcome. Grammar attendance alone increases the probability of achieving A-levels by 26 percentage points ($p < 0.05$) in the group of compliers. The effect of grammar is significant and positive for obtaining a university degree only with conventional standard errors (not shown, $p < 0.05$), while the larger bootstrapped standard errors return a smaller t-test result, indicating a statistically in-

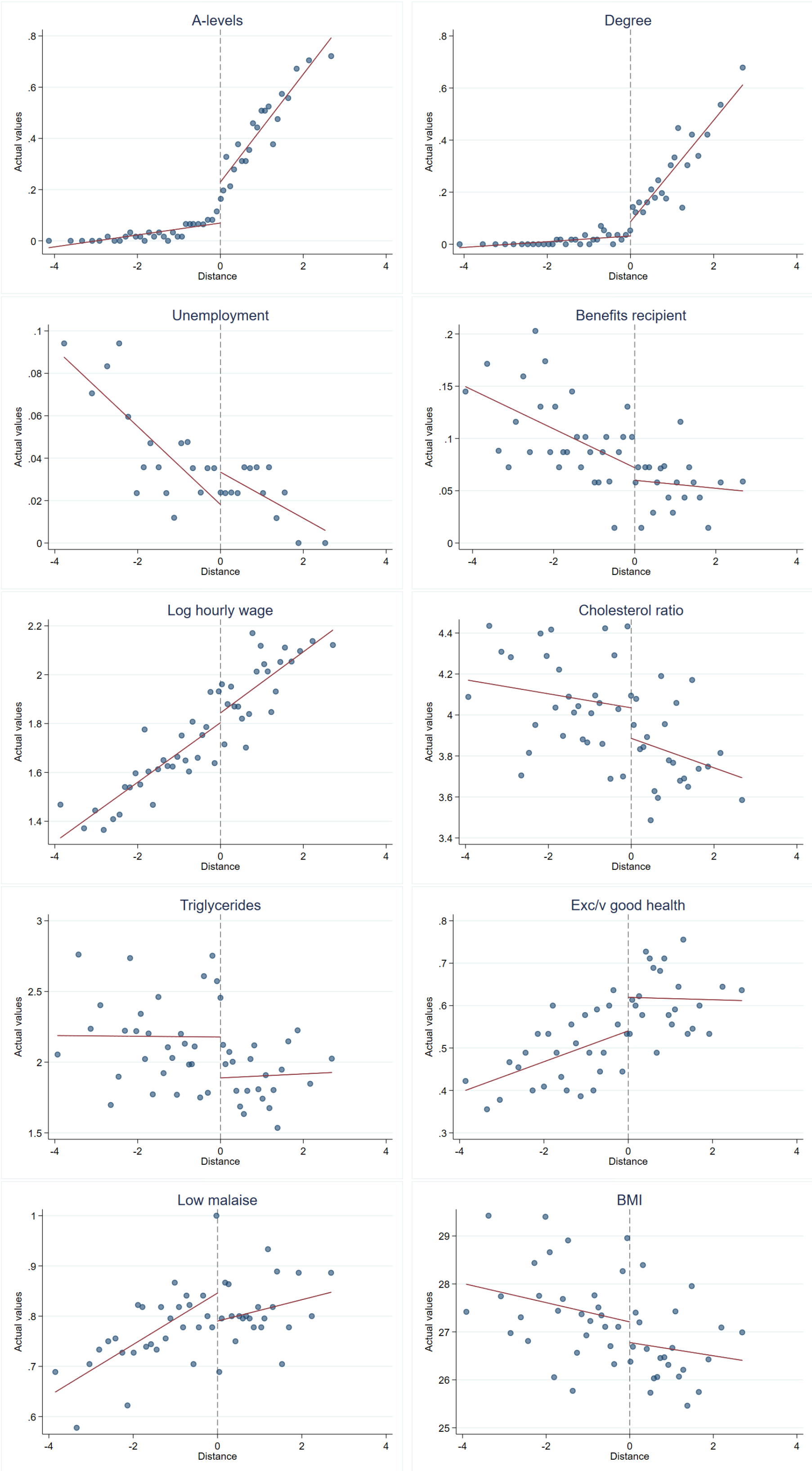

Figure 3: Outcomes as a function of the distance variable R_{LEA}.

significant effect ($t = 1.22$, $p = 1.776$). For both A-levels and obtaining a university degree however, the interaction variable between distance and threshold indicator is positive and significant, indicating a positive change in the slope of the probability of both outcomes as a function of the ability score after the grammar entry threshold. Considering grammar schools were the track with highest academic focus, this result is not surprising, and it is clearly reflected in the figures.

On the other hand, the estimated effects of grammar for adult labour, health and disease risk outcomes are not significant in the complier group. The interaction between distance score and the threshold dummy is significant for BMI and triglycerides level. The effect is large and negative for BMI, indicating lower average adult BMI after the grammar entry threshold. Conversely, it is positive for triglycerides, although this is offset by the negative and significant effect of ability score distance from the threshold. As anticipated by the OLS regression tables, other background covariates play a prominent role in determining a variety of human capital outcomes. Non-cognitive skills, sex, mother's interest in child education, father's SES and child morbidity index are all significantly associated with some if not most outcomes. Note that distance (cognitive score rescaled by LEA cutoff) is not particularly significant in most cases, likely because it is more homogeneous once we select observations within the bandwidths on both sides of the threshold. Moreover, selection effects appear to originate early in utero, supporting the fetal origins literature: higher frequency of smoking during pregnancy by the mother shows a positive association with unemployment in adulthood and triglycerides levels. On the other hand however, effects are mixed, since smoking during pregnancy is also associated with lower probability of being a benefits recipient, lower malaise and lower cholesterol ratio levels. Finally, higher childhood morbidity is associated with lower probability of obtaining A-levels and a degree, higher probability of unemployment and lower hourly wages at 33, as well as higher BMI and triglycerides.

Table 6 displays results of the fuzzy RDD estimation, implementing MSE-optimal bandwidths with the bias-correction procedure proposed by Calonico, Cattaneo, Farrell, and Titiunik (2017) and Calonico, Cattaneo, and Titiunik (2014a). Results are fairly consistent with those estimated without bias correction. The samples obtained with bias-corrected MSE-optimal bandwidths are generally smaller, except for hourly wage, cholesterol ratio and triglyceride levels, where samples are slightly larger. When using standard errors correcting for bias in the bandwidth and treatment effect estimate, grammar attendance is significant for achieving any A-levels and a university degree ($p < 0.01$ and $p < 0.05$ respectively), and magnitudes are slightly smaller. Health outcomes and risk of disease are unaffected by grammar attendance, with the exception of the indicator for low malaise score, negatively affected by grammar attendance. Contrary to the education outcomes however, this result was not robust to varying choice of polynomial and bandwidth. Notice also that discontinuity estimates in the grammar assignment function (first-stage estimates) are all highly significant, offering support to the validity of the study design.

4.4 Robustness checks

As a placebo test for the first stage, we run all outcome regressions with bias correction at a fictional cutoff of 0.2[11]. The aim is to check that by estimating regressions using 0.2 as cutoff for treatment assignment we do not get significant first and second stages. Table 7 shows that treatment effect estimate is not significant for any outcomes. Moreover first-stage estimates are not significant either, meaning no discontinuities are detected at 0.2 in the distance variable. Figure 4, plotting the probability of grammar against ability scores with different thresholds, also shows that the discontinuity in treatment assignment is not detected once we move away from the threshold used in the main specification.

As a second robustness check we re-estimate the model with the donut exclusion around the threshold. The density function for distance from LEA-specific threshold in Figure 5 presents a concentration of observations around the threshold, and we aim at showing that our results are not sensitive to excluding these. MSE-optimal bandwidth selection is operated again after excluding the observations within 0.8 from the threshold, dropping 4% of observations (n=148). Both the specification with and without bias correction confirm the main results: A-levels and obtaining a degree are positively and significantly associated with grammar attendance, and coefficients are slightly larger in magnitude, which we would expect since individuals become less similar as we move away from the threshold (see Table 8 in main text and Tables A5-A6 in the Appendix). Confirming previous results, grammar is not significantly associated with labour market and health outcomes. An exception is represented by the dummy for benefits recipients, which is negatively associated to grammar when excluding observations closest to the threshold, although this effect is partly offset by the positive and significant coefficients of the distance variable and its interaction with the threshold dummy.

Finally, we test our approach on maths scores and BMI at age 7, in the same spirit as Manning and Pischke (2006), who criticise value-added approaches to estimate the effect of grammar school attendance. We would expect outcomes prior to secondary school not to be affected by attendance to a particular type of secondary school if our approach successfully isolates treatment effect from other confounders. Note that Manning and Pischke (2006) implemented their check on age 11 outcomes, while we prefer age 7 outcomes since we use age 11 maths, reading and general ability scores to derive our age 11 cognitive ability index. Since the index constitutes the running variable for our RDD, we would expect a significant relationship with age 11 maths scores purely because of how we construct our variable. We find that age 7 maths scores are not affected by the

[11] Placebo tests executed without correcting for bias are presented in the Appendix

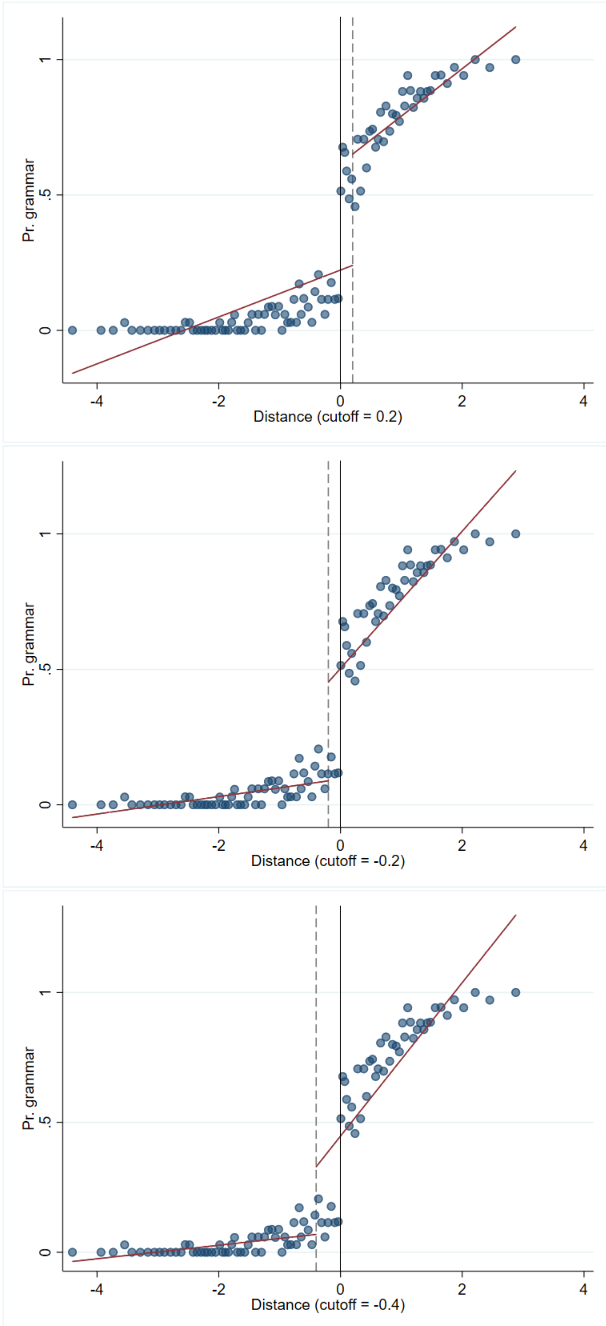

Figure 4: Probability of grammar as a function of distance from the LEA-specific threshold, with placebo cutoffs at 0.2, -0.2 and -0.4. Note that the scatter points show a sharp discontinuity at 0 and that any other cutoff is unable to account for that.

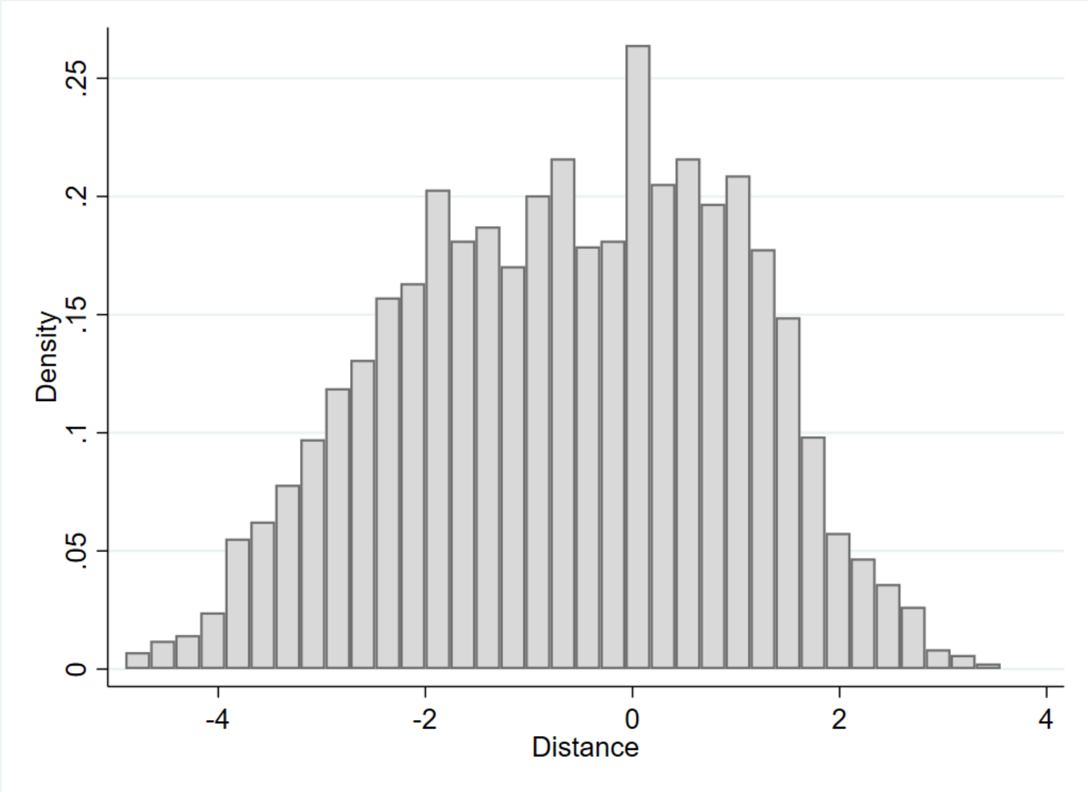

Figure 5: Histogram illustrating density of distance from LEA-specific threshold, $A_i - c_{LEA}$.

discontinuity in grammar attendance, while age 16 scores are, a finding that holds both with and without bias correction, as well as in the reduced form equation (see Table 9). This result further increases our confidence that the identification strategy is correctly isolating the effect of grammar attendance. Neither age 7 nor age 16 BMI is affected by the discontinuity, a finding that is not surprising given that this and other health outcomes were not found to be affected in our main specifications either.

4.5 Results of the mechanism analysis

We implement the RDD approach to investigate the effect of grammar attendance on a range of school quality indicators that could serve as channels between grammar school attendance and accumulation of human capital, as detailed in Section 3.3. Summary characteristics highlighting differences by school type are shown in Table 10. For simplicity we dichotomise all mechanism variables when it comes to treating them as the endogenous variable in the first stage of the fuzzy RDD. Grammar pupils are almost twice as likely to attend schools with higher socioeconomic status pupils: these are schools where 50% or more of pupils' fathers are in non-manual occupations, and schools that prior to the 1972 reform had high attendance rates beyond compulsory school-leaving age. Grammar pupils are twice as likely to attend a school with a high (above NCDS median) share of girls taking only full GCE qualifications, indicative of higher pupil ability[12]. On the other hand, secondary modern pupils are twice as likely as grammar ones to attend a school displaying a high (above NCDS median) proportion of girls taking only the partial qualification, leading to a CSE. Finally, similar proportions of grammar and secondary modern pupils attend schools lacking a library (about 1 in 4), while secondary modern pupils are more likely to lack science or sports facilities at school.

Table 11 shows results for mechanism analysis, with MSE-optimal bandwidths and bias correction. Being above the cutoff for grammar is used as instrument to predict each of the mechanisms in turn, and their respective effect on human capital outcomes. The only mechanisms that produce any effect on outcomes are those related to pupil ability. Attending a school with above-median proportion of girls attending the full qualification course (GCE) increases individual probability of achieving any A-levels by 50 percentage points and a degree by 38 percentage points ($p < 0.05$). On the other hand having an above-median proportion of girl pupils taking the lower qualification, CSE, not requiring completion of the school cycle, is negatively associated with both A-levels and degree ($p < 0.1$). Marginally significant effects are also found for the probability of being unemployed at age 33 ($p < 0.1$): this is lower for higher proportions of girls taking GCEs, and higher for higher proportion of girl CSE-takers. The mechanism analysis therefore points to an effect of peer ability on education outcomes, but does not highlight any role of socioeconomic status of school pupils nor of school resources. Unfortunately not much information was available on the quality of teaching and school curriculum, which are also possible mechanisms for the effect of grammar schools.

5 Discussion and concluding remarks

Results indicate that when isolating its effect from confounders, grammar attendance is a significant predictor of education outcomes, but not of labour market outcomes, health or risk of developing illness. We would expect grammar attendance to be a significant and large predictor of A-levels obtained and having a university degree, since secondary modern schools at the time were geared towards vocational professions, and only very rarely their pupils completed any A-level or degree qualifications. This point deserves our attention. Examining the effect of grammar attendance for a cohort born in 1958 today allows us to explore individuals' life trajectories over an unusually long period of time, but this also means that the school landscape has changed quite considerably since. In particular, obtaining A-levels and a university degree is now much more common than it used to be, and grammar schools are not the only public institutions offering a

[12]The reason why we took share of girls rather than total share of pupils was that figures for shares of boys and girls were provided separately in the NCDS, and there was no reliable way to calculate the total share of pupils taking GCEs or CSEs.

more academic education. This means that the grammar advantage in educational terms observed in the 1958 generation may not apply for current young generations of 11-plus takers.

However, as already noted, most studies analysing different contexts and cohorts find significant and positive effects of higher quality schools on educational outcomes, while not as often for labour market or health outcomes. In their study of the German context, Dustmann et al. (2016) link the lack of an effect of track assignment on education and labour market outcomes to the possibility of switching tracks in later grades. In the English system this was rarely the case, but we similarly find that grammar school attendance alone cannot explain differences in most individual outcomes, except for those directly related to educational achievements. Similar results are also found in the British context by Clark (2010) and Del Bono and Clark (2016), in the smaller areas of Yorkshire and Aberdeen respectively. Importantly, the present treatment effect estimates hold for individuals who got the marginal place available in grammar schools. However, a recent paper by Janke et al. (2018), examining the long-term effect of reforms increasing the quantity of schooling in the UK, similarly concludes that they had no significant effect on health along the whole educational attainment distribution. Finally, our results are also in line with a recent paper by Courtin et al. (2019) on the effect of an increase in compulsory schooling in France. Looking at quantity, rather than quality, of schooling, they find educational effects only, and no effect on biomarkers. Thus, observable differences in health and risk of disease among individuals who attended different schools is to be linked to earlier factors, such as cognitive and non-cognitive skills prior to secondary school, which remain highly predictive of human capital outcomes across all specifications.

We have provided an empirical investigation of the long-term effects of grammar school attendance on human capital with a quasi-experimental methodology, exploiting a discontinuity in the probability of admission and building a novel strategy for threshold estimation from limited information. We offer a contribution to the body of research informing educational policy-makers on the effects of selective schools as means to tailor school quality to student ability. We conclude that the marginal student admitted to grammar school does not benefit in terms of long-term human capital accumulation, with the exception of the direct positive effect on education outcomes. A more prominent role might be played by the child's cognitive and non-cognitive skills, as well as parental support. Further research could help assess the overall impact of the grammar school system on pupils of all abilities, but we anticipate that the large role played by background characteristics will persist.

References

Abdulkadiroglu, A., Angrist, J. D., and Pathak, P. A. (2014). 'The Elite Illusion: Achievement Effects at Boston and New York Exam Schools'. *Econometrica* vol. 82 (1), pp. 137–196.

Adams, R. (May 11, 2018). 'Grammar schools in England to get £50m expansion fund'. *The Guardian.* URL: `https://www.theguardian.com/education/2018/may/11/grammar-schools-in-england-to-get-50m-expansion-fund` (visited on 05/17/2018).

Atkinson, A., Gregg, P., and McConnell, B. (2006). 'The Result of 11 Plus Selection: An Investigation into Opportunities and Outcomes for Pupils in Selective LEAs'. Centre for Market and Public Organisation Working Paper Series No. 06/150.

Bai, J. (1997). 'Estimation of a change point in multiple regression models'. *Review of Economics and Statistics* vol. 79 (4), pp. 551–563.

Bai, J., Lumsdaine, R. L., and Stock, J. H. (1998). 'Testing For and Dating Common Breaks in Multivariate Time Series'. *Review of Economic Studies* vol. 65, pp. 395–432.

Barreca, A. I., Lindo, J. M., and Waddell, G. R. (2016). 'Heaping-induced bias in regression-discontinuity designs'. *Economic Inquiry* vol. 54 (1), pp. 268–293.

Basu, A., Jones, A. M., and Rosa Dias, P. (2018). 'Heterogeneity in the impact of type of schooling on adult health and lifestyle'. *Journal of Health Economics* vol. 57, pp. 1–14.

Benzeval, M., Davillas, A., Kumari, M., and Lynn, P. (2014). 'Biomarker User Guide and Glossary'. URL: `https://www.understandingsociety.ac.uk/sites/default/files/downloads/legacy/7251-UnderstandingSociety-Biomarker-UserGuide-2014-1.pdf` (visited on 09/21/2017).

Bonhomme, S. and Sauder, U. (2011). 'Recovering distributions in difference-in-differences models : a comparison of selective and comprehensive schooling'. *Review of Economic Studies* vol. 93, pp. 479–494.

Brown, M., Dodgeon, B., and Mostafa, T. (2016). *Webinar: Introduction to the National Child Development Study.* URL: `http://www.cls.ioe.ac.uk/library-media/documents/NCDS%20Webinar%20May%202016.pdf` (visited on 01/17/2018).

Burgess, S., Dickson, M., and Macmillan, L. (2019). 'Do selective schooling systems increase inequality?' *Oxford Economic Papers*, pp. 1–24.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2017). 'rdrobust: Software for regression-discontinuity designs'. *The Stata Journal* vol. 17 (2), pp. 372–404.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014a). 'Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs'. *Econometrica* vol. 82 (6), pp. 2295–2326.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018). ‘Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs’. Unpublished manuscript. URL: https://arxiv.org/pdf/1809.00236.pdf.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2018). ‘Regression Discontinuity Designs Using Covariates’. *Review of Economics and Statistics (forthcoming).*

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014b). ‘Robust data-driven inference in the regression-discontinuity design’. *The Stata Journal* vol. 14 (4), pp. 909–946.

Card, D. and Giuliano, L. (2016). ‘Can Tracking Raise the Test Scores of High-Ability Minority Students?’ *American Economic Review* vol. 106 (10), pp. 2783–2816.

Card, D., Mas, A., and Rothstein, J. (2008). ‘Dynamic Models of Segregation’. *The Quarterly Journal of Economics* vol. 123 (1), pp. 177–218.

Case, A., Fertig, A., and Paxson, C. (2005). ‘The lasting impact of childhood health and circumstance’. *Journal of Health Economics* vol. 24 (2), pp. 365–389.

Cattaneo, M. D. and Vazquez-Bare, G. (2016). ‘The Choice of Neighborhood in Regression Discontinuity Designs’. *Observational Studies* vol. 2, pp. 134–146.

Cawley, J., Conneely, K., Heckman, J., and Vytlacil, E. (1997). ‘Cognitive Ability, Wages, and Meritocracy’. *Intelligence, Genes, and Success: Scientists Respond to The Bell Curve.* Ed. by B. Devlin, S. E. Fienberg, D. P. Resnick, and K. Roeder. New York, NY: Springer New York, pp. 179–192.

Clark, D. (2010). ‘Selective Schools and Academic Achievement’. *The B.E. Journal of Economic Analysis & Policy* vol. 10 (1). Article 9.

Courtin, E., Nafilyan, V., Avendano, M., Meneton, P., Berkman, L. F., Goldberg, M., Zins, M., and Dowd, J. B. (2019). ‘Longer schooling but not better off? A quasi-experimental study of the effect of compulsory schooling on biomarkers in France’. *Social Science & Medicine* vol. 220, pp. 379–386.

Dearden, L., Ferri, J., and Meghir, C. (2002). ‘The Effect of School Quality on Educational Attainment and Wages’. *The Review of Economics and Statistics* vol. 84 (1), pp. 1–20.

Del Bono, E. and Clark, D. (2016). ‘The Long-Run Effects of Attending an Elite School: Evidence from the UK’. *American Economic Journal: Applied Economics* vol. 8 (1), pp. 150–176.

Dobbie, W. and Fryer, R. G. (2014). ‘The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools’. *American Economic Journal: Applied Economics* vol. 6 (3), pp. 58–75.

Dong, Y. (2018). ‘Jump or Kink? Regression Probability Jump and Kink Design for Treatment Effect Evaluation’. Unpublished manuscript. URL: http://www.yingyingdong.com/Research/RDwithoutDiscontinuity.pdf.

Dustan, A. (2010). ‘Have Elite Schools Earned their Reputation?: High School Quality and Student Tracking in Mexico City’. Unpublished manuscript. URL: https://are.berkeley.edu/fields/ae/dustan.pdf.

Dustmann, C., Puhani, P. A., and Schönberg, U. (2016). ‘The Long-term Effects of Early Track Choice’. *Economic Journal* vol. 127, pp. 1348–1380.

Galindo-Rueda, F. and Vignoles, A. (2005). ‘The Heterogeneous Effect of Selection in Secondary Schools: Understanding the Changing Role of Ability.’ Working Paper No. 52, LSE Centre for the Economics of Education.

Gelman, A. and Imbens, G. (2018). ‘Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs’. *Journal of Business and Economic Statistics (forthcoming)*. URL: http://www.tandfonline.com/action/journalInformation?journalCode=ubes20.

Guyon, N., Maurin, E., and McNally, S. (2012). ‘The Effect of Tracking Students by Ability into Different Schools: A Natural Experiment’. *Journal of Human Resources* vol. 47 (3), pp. 684–721.

Hahn, J., Todd, P. E., and Van der Klaauw, W. (2001). ‘Notes and Comments: Identification and estimation of treatment effects with a regression-discontinuity design.’ *Econometrica* vol. 69 (1), pp. 201–209.

Hall, C. (2012). ‘Does making upper secondary school more comprehensive affect dropout rates, educational attainment and earnings? Evidence from a Swedish pilot scheme’. *Journal of Human Resources* vol. 47 (1), pp. 237–269.

Harmon, C. and Walker, I. (2000). ‘The Returns to the Quantity and Quality of Education: Evidence for Men in England and Wales’. *Economica* vol. 67 (265), pp. 19–35.

Imbens, G. and Kalyanaraman, K. (2012). ‘Optimal Bandwidth Choice for the Regression Discontinuity Estimator’. *The Review of Economic Studies* vol. 79 (3), pp. 933–959.

Janke, K., Johnston, D. W., Propper, C., and Shields, M. A. (2018). *The Causal Effect of Education on Chronic Health Conditions*. Tech. rep. IZA Discussion Paper No. 11353.

Jeffreys, B. (Aug. 1, 2018). ‘Grammar schools: Thousands of new places created’. *BBC News*. URL: www.bbc.co.uk/news/education-44727857 (visited on 08/02/2018).

Jones, A. M., Rice, N., and Rosa Dias, P. (2011). ‘Long-Term Effects of School Quality on Health and Lifestyle : Evidence from Comprehensive Schooling Reforms in England’. *Journal of Human Capital* vol. 5 (3), pp. 342–376.

Jones, A. M., Rice, N., and Rosa Dias, P. (2012). ‘Quality of schooling and inequality of opportunity in health’. *Empirical Economics* vol. 42, pp. 369–394.

Kautz, T., Heckman, J. J., Diris, R., Weel, B. T., and Borghans, L. (2014). ‘Fostering and Measuring Skills : Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success’. OECD Education Working Paper No. 110.

Kerckhoff, A. (1986). 'Efffects of Ability Grouping in Secondary Schools in Great Britain'. *American Sociological Review* vol. 51 (6), pp. 842–858.

Kerr, S., Pekkarinen, T., and Uusitalo, R. (2013). 'School Tracking and Development of Cognitive Skills'. *Journal of Labour Economics* vol. 31 (3), pp. 577–602.

Kirabo Jackson, C. (2010). 'Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago'. *Economic Journal* vol. 120 (549), pp. 1399–1429.

Lee, D. S. and Lemieux, T. (2010). 'Regression Discontinuity Designs in Economics'. *Journal of Economic Literature* vol. 48, pp. 281–355.

Manning, A. and Pischke, J. S. (2006). 'Comprehensive Versus Selective Schooling in England in Wales: What We Know?' NBER Working Paper No. 12176.

McCrary, J. (2008). 'Manipulation of the running variable in the regression discontinuity design: A density test'. *Journal of Econometrics* vol. 142, pp. 698–714.

Pop-Eleches, C. and Urquiola, M. (2013). 'Going to a Better School: Effects and Behavioral Responses'. *American Economic Review* vol. 103 (4), pp. 1289–1324.

Porter, J. and Yu, P. (2015). 'Regression discontinuity designs with unknown discontinuity points: Testing and estimation '. *Journal of Econometrics* vol. 189, pp. 132–147. URL: `http://dx.doi.org/10.1016/j.jeconom.2015.06.002`.

University of London. Institute of Education. Centre for Longitudinal Studies. (2008a). *National Child Development Study: Sweep 4, 1981, and Public Examination Results, 1978.* [data collection]. 2nd Edition. National Children's Bureau, [original data producer(s)]. UK Data Service. SN: 5566. URL: `http://doi.org/10.5255/UKDA-SN-5566-1`.

— (2008b). *National Child Development Study: Sweep 5, 1991.* [data collection]. 2nd Edition. City University. Social Statistics Research Unit, [original data producer(s)]. UK Data Service. SN: 5567. URL: `http://doi.org/10.5255/UKDA-SN-5567-1`.

— (2008c). *National Child Development Study: Sweep 6, 1999-2000.* [data collection]. 2nd Edition. Joint Centre for Longitudinal Research, [original data producer(s)]. UK Data Service. SN: 5578. URL: `http://doi.org/10.5255/UKDA-SN-5578-1`.

— (2012). 'National Child Development Study: Sweep 8, 2008-2009.' [data collection]. 3rd Edition. UK Data Service. SN: 6137. URL: `http://doi.org/10.5255/UKDA-SN-6137-2`.

— (2014). *National Child Development Study: Childhood Data, Sweeps 0-3, 1958-1974.* [data collection]. 3rd Edition. National Birthday Trust Fund, National Children's Bureau, [original data producer(s)]. UK Data Service. SN: 5565. URL: `http://doi.org/10.5255/UKDA-SN-5565-2`.

Tables

Table 1: Summary of related literature.

	Country	Outcomes			Method
		Educ.	Labour	Health	
Abdulkadiroglu et al. (2014)	United States	x			RDD
Atkinson et al. (2006)	England	x			Matching and logistic regressions
Basu et al. (2018)	England			x	IV
Bonhomme and Sauder (2011)	England	x			VAR and DID
Burgess et al. (2019)	England		x		Matching
Clark (2010)	England	x			RDD
Del Bono and Clark (2016)	Scotland	x	x	Fertility	RDD
Dobbie and Fryer (2014)	United States	x			RDD
Dustan (2010)**	Mexico	x			RDD
Dustmann et al. (2016)	Germany	x	x		RDD
Galindo-Rueda and Vignoles (2005)*	England	x			VAR and IV
Guyon et al. (2012)	Northern Ireland	x			RDD
Hall (2012)	Sweden	x	x		IV
Harmon and Walker (2000)	England		x		OLS and IV
Jones, Rice, and Rosa Dias (2011)	England			x	Matching and OLS
Jones, Rice, and Rosa Dias (2012)	England			x	FSD and distributional regressions
Kerckhoff (1986)	England	x			OLS regressions
Kerr et al. (2013)	Finland	Skills			DID
Kirabo Jackson (2010)	Trinidad and Tobago	x			RDD
Manning and Pischke (2006)*	England	x			VAR, IV and placebo tests
Pop-Eleches and Urquiola (2013)	Romania	x			RDD

Unpublished working papers (*) and PhD paper (**).

Value-added regressions (VAR), instrumental variables (IV), difference-in-differences (DID), first-order stochastic dominance (FSD).

Table 2: Descriptive statistics of childhood characteristics and outcomes by type of secondary school attended

	Grammar				Secondary modern			
	Mean	s.d.	Min	Max	Mean	s.d.	Min	Max
Baseline characteristics								
Cognitive skills	1.80	0.83	-3	4	-0.51	1.27	-4	3
Non-cognitive skills	0.94	0.08	0	1	0.86	0.13	0	1
Female	0.55	0.50	0	1	0.49	0.50	0	1
Mother's interest	2.70	0.77	0	4	1.88	1.03	0	4
Father's SES	3.32	0.90	1	5	2.82	0.81	1	5
Mother smoke preg.	1.37	0.78	1	4	1.58	0.92	1	4
Child morbidity	0.06	0.03	0	0	0.06	0.04	0	0
Outcomes								
Educational attainment								
Any A-levels	0.51	0.50	0	1	0.03	0.16	0	1
University degree	0.31	0.46	0	1	0.02	0.14	0	1
Labour market (age 33)								
Unemployment	0.02	0.15	0	1	0.04	0.21	0	1
Benefits recipient	0.05	0.22	0	1	0.10	0.30	0	1
Hourly wage at 33	9.35	12.27	0	148	6.57	11.61	0	109
Self-assessed health (age 50)								
Exc./very good health	0.62	0.49	0	1	0.50	0.50	0	1
Low malaise	0.81	0.39	0	1	0.77	0.42	0	1
Health visits (age 45)								
Body Mass Index (BMI)	26.40	4.68	17	51	27.56	5.04	18	52
Cholesterol ratio	3.80	1.14	2	8	4.07	1.18	2	12
Triglycerides	1.87	1.45	0	17	2.17	1.75	0	27
Observations	1160				2288			

Table 3: OLS regressions for whole sample.

	A-levels		Degree		Unemployed		On benefits		Log wage	
Grammar	0.4876***	0.3467***	0.2955***	0.1729***	-0.0163*	0.0147	-0.0476***	-0.0068	0.3300***	0.0600
	(0.0137)	(0.0190)	(0.0130)	(0.0179)	(0.0079)	(0.0112)	(0.0106)	(0.0149)	(0.0316)	(0.0403)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
F Stat.	1263.121	187.235	519.282	83.124	4.220	4.823	20.056	8.943	108.920	68.119
Observations	2505	2505	2352	2352	2116	2116	2816	2816	1553	1553
	High SAH		Low malaise		BMI		Chol		Trig	
Grammar	0.1238***	0.0197	0.0278	-0.0401	-1.2002***	-0.4132	-0.2549***	-0.0565	-0.2727***	-0.1288
	(0.0235)	(0.0329)	(0.0194)	(0.0270)	(0.2393)	(0.3330)	(0.0621)	(0.0813)	(0.0801)	(0.1081)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
F Stat.	27.719	8.806	2.047	8.904	25.160	9.958	16.826	38.082	11.584	23.945
Observations	1869	1869	1865	1865	1786	1786	1498	1498	1500	1500

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Human capital outcomes: polynomial regressions for sample with pre-selected bandwidth with bootrapped standard errors. Running variable is the distance between cognitive score and threshold, which is LEA-specific.

	(1)	(2)	(3)	(4)	(5)
	A-levels	Degree	Unemployed	On benefits	Log wage
Grammar	0.261*	0.146	0.0027	-0.0454	0.0693
	(0.127)	(0.119)	(0.0503)	(0.0801)	(0.237)
Distance	0.0175	-0.0132	0.0052	0.0090	0.0863
	(0.0595)	(0.0572)	(0.0290)	(0.0472)	(0.120)
Distance $\times 1[A \geq c_{LEA}]$	0.136*	0.140**	-0.0007	0.0085	-0.0127
	(0.0601)	(0.0517)	(0.0313)	(0.0429)	(0.108)
Non-cognitive skills	0.0954	0.111	-0.148**	-0.191**	0.313+
	(0.0769)	(0.0676)	(0.0447)	(0.0610)	(0.158)
Female	-0.0316*	-0.0266*	-0.0007	0.0477***	-0.471***
	(0.0134)	(0.0113)	(0.0062)	(0.0084)	(0.0217)
Mother's interest	0.0211*	0.0118	0.0018	-0.0076	0.0399+
	(0.0094)	(0.00731)	(0.0040)	(0.0059)	(0.0216)
Father's SES	0.0785***	0.0563***	0.0019	-0.0091	0.0860***
	(0.0081)	(0.00846)	(0.0040)	(0.0061)	(0.0136)
Mother smoke preg.	-0.0011	-0.00288	0.0096**	-0.0065*	-0.0131
	(0.0051)	(0.00452)	(0.0028)	(0.0032)	(0.0111)
Child morbidity	-0.323*	-0.436***	0.269***	-0.0549	-1.123***
	(0.131)	(0.116)	(0.0683)	(0.0848)	(0.257)
Constant	-0.277**	-0.239**	0.125*	0.292***	1.494***
	(0.0975)	(0.0877)	(0.0567)	(0.0630)	(0.229)
First-stage $1[A \geq c_{LEA}]$	0.4637***	0.4535***	0.4361***	0.4444***	0.4507***
	(0.0340)	(0.0344)	(0.0402)	(0.0310)	(0.0514)
First-stage F	185.989	174.073	117.957	205.661	76.830
Observations in bandwidth	1599	1538	1213	1849	791
Total observations available	2505	2352	2116	2816	1553

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Health outcomes: polynomial regressions for sample with pre-selected bandwidth with bootrapped standard errors. Running variable is the distance between cognitive score and threshold, which is LEA-specific.

	(6) High SAH	(7) Low malaise	(8) BMI	(9) Chol	(10) Trig
Grammar	0.0454	-0.0750	-0.571	-0.398	1.020
	(0.190)	(0.117)	(1.735)	(0.392)	(0.653)
Distance	0.0556	0.0329	1.466	0.203	-1.295**
	(0.0988)	(0.0696)	(1.075)	(0.208)	(0.360)
Distance $\times 1[A \geq c_{LEA}]$	-0.00267	-0.0502	-2.679*	-0.337	1.116**
	(0.105)	(0.0743)	(1.186)	(0.234)	(0.302)
Non-cognitive skills	0.185	0.107	-1.222	-0.0814	3.074***
	(0.136)	(0.104)	(1.086)	(0.227)	(0.300)
Female	0.0392+	-0.115***	-0.832***	-0.970***	-0.485***
	(0.0210)	(0.0128)	(0.165)	(0.0348)	(0.0506)
Mother's interest	0.0306*	0.00809	-0.249	-0.00497	-0.106
	(0.0137)	(0.0100)	(0.183)	(0.0394)	(0.0793)
Father's SES	0.0653***	0.0139	-0.896***	-0.0731*	-0.433***
	(0.0153)	(0.0101)	(0.104)	(0.0340)	(0.0445)
Mother smoke preg.	-0.0152+	-0.0233***	-0.00396	-0.0739*	0.289***
	(0.00901)	(0.00613)	(0.0672)	(0.0325)	(0.0196)
Child morbidity	-0.212	0.312	6.741**	0.477	6.805***
	(0.194)	(0.208)	(1.990)	(0.557)	(0.589)
Constant	0.111	0.766***	32.29***	5.104***	-0.343
	(0.204)	(0.153)	(1.995)	(0.504)	(2.313)
First-stage $1[A \geq c_{LEA}]$	0.4692***	0.4608***	0.4717***	0.4434***	0.4388***
	(0.0550)	(0.0473)	(0.0574)	(0.0617)	(0.0623)
First-stage F	72.710	95.030	67.538	51.711	50.402
Observations in bandwidth	771	974	716	605	607
Total observations available	1869	1865	1786	1498	1500

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Bias-corrected model implementing fuzzy RDD implemented using distance between score and threshold as the assignment variable. Thresholds are imputed from the data and LEA-specific. Estimates are from rdrobust command in Stata(by Calonico et al. 2014, 2017). First stage estimate and conventional s.e. are included.

	(1) A-levels	(2) Degree	(3) Unemployed	(4) On benefits	(5) Log gross income
RD_Estimate	0.2458**	0.1773*	-0.0245	-0.0644	0.1317
	(0.0761)	(0.0797)	(0.0526)	(0.0519)	(0.1391)
Robust 95% CI	[.038 , .401]	[.04 , .411]	[-.147 , .096]	[-.223 , .024]	[-.192 , .502]
Bandwidth	1.3036	0.6716	0.8144	1.5038	1.4889
Left of c	622	315	357	818	454
Right of c	741	397	417	916	533
Available obs.	2505	2352	2116	2816	1553
First-stage estimate	0.4579***	0.4592***	0.4640***	0.4492***	0.4586***
First-stage conv. s.e.	(0.0405)	(0.0551)	(0.0523)	(0.0351)	(0.0464)
	(6)	(7)	(8)	(9)	(10)
	SAH	Malaise	BMI	Chol	Trig
RD_Estimate	0.0574	-0.2175*	-1.0928	-0.4546	-0.5793
	(0.1464)	(0.1032)	(1.4770)	(0.3606)	(0.5235)
Robust 95% CI	[-.312 , .374]	[-.519 , -.036]	[-4.633 , 2.121]	[-1.231 , .515]	[-1.753 , .901]
Bandwidth	0.7865	0.6958	0.7834	1.1720	1.2202
Left of c	301	261	292	360	379
Right of c	380	339	370	430	452
Available obs.	1869	1865	1786	1498	1500
First-stage estimate	0.4936***	0.5218***	0.4835***	0.4240***	0.4320***
First-stage conv. s.e.	(0.0563)	(0.0592)	(0.0553)	(0.0509)	(0.0496)

Standard errors in parentheses

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

Table 7: Fuzzy RDD implemented as a falsification test at a threshold 0.2. Estimates are from rdrobust command in Stata(by Calonico et al. 2014, 2017). First stage estimate and conventional s.e. are included.

	(1)	(2)	(3)	(4)	(5)
	A-levels	Degree	On benefits	Unemployed	Log hrly wage
RD_Estimate	-0.0816	-0.3967	-0.3376	-0.1413	-1.5108
	(0.5100)	(0.5821)	(0.2803)	(0.3020)	(1.5986)
Robust 95% CI	[-1.092 , 1.016]	[-1.563 , .92]	[-.89 , .268]	[-.762 , .507]	[-4.667 , 2.021]
Bandwidth	0.9349	0.9147	0.9618	0.8183	0.9070
Left of c	491	482	585	376	310
Right of c	523	467	582	378	326
Available obs.	2505	2352	2816	2116	1553
First-stage estimate	-0.1577	-0.1386	-0.1509	-0.1169	-0.1267
First-stage conv. s.e.	(0.0818)	(0.0865)	(0.0762)	(0.0945)	(0.1039)
	(6)	(7)	(8)	(9)	(10)
	SAH	Malaise	BMI	Chol	Trig
RD_Estimate	-0.2315	-0.2373	-11.5633	3.4154	1.9289
	(0.7876)	(0.3950)	(11.7821)	(6.5251)	(3.2341)
Robust 95% CI	[-1.829 , 1.441]	[-1.105 , .544]	[-35.893 , 13.976]	[-12.128 , 15.272]	[-5.511 , 8.091]
Bandwidth	1.0325	0.6922	0.8501	1.0942	0.9881
Left of c	419	292	342	370	343
Right of c	463	315	365	373	347
Available obs.	1869	1865	1786	1498	1500
First-stage estimate	-0.1256	-0.2219+	-0.1345	-0.0570	-0.1024
First-stage conv. s.e.	(0.0886)	(0.1083)	(0.1039)	(0.0934)	(0.0987)

Standard errors in parentheses

$+p < 0.1$

Table 8: Fuzzy RDD implemented using distance between score and threshold as the assignment variable, with donut exclusion around the threshold. Thresholds are imputed from the data and LEA-specific. Estimates are from rdrobust command in Stata(by Calonico et al. 2014, 2017). First stage estimate and conventional s.e. are included.

	(1)	(2)	(3)	(4)	(5)
	A-levels	Degree	Unemployed	On benefits	Log gross income
RD_Estimate	0.2971**	0.2549***	0.0459	-0.2032*	0.1986
	(0.0942)	(0.0730)	(0.0471)	(0.0962)	(0.1922)
Robust 95% CI	[.079 , .522]	[.114 , .454]	[-.076 , .148]	[-.478 , -.028]	[-.138 , .764]
Bandwidth	1.0844	0.9915	1.0621	0.8253	0.9329
Left of c	454	396	405	365	245
Right of c	611	525	515	532	351
Total obs. available	2393	2246	2020	2691	1481
First-stage estimate	0.4603***	0.4646***	0.4631***	0.4565***	0.4802***
First-stage conv. s.e.	(0.0476)	(0.0526)	(0.0521)	(0.0526)	(0.0638)
	(6)	(7)	(8)	(9)	(10)
	SAH	Malaise	BMI	Chol	Trig
RD_Estimate	0.2648+	0.1737	-1.9302	-0.3216	-0.4039
	(0.1569)	(0.1300)	(1.6181)	(0.3737)	(0.4772)
Robust 95% CI	[-.146 , .604]	[-.16 , .475]	[-6.172 , 1.741]	[-1.288 , .531]	[-1.577 , .653]
Bandwidth	1.0875	1.1027	1.1017	1.0264	0.8518
Left of c	365	366	355	268	210
Right of c	486	491	460	363	307
Total obs. available	1778	1774	1696	1421	1422
First-stage estimate	0.4605***	0.4584***	0.4673***	0.5030***	0.5245***
First-stage conv. s.e.	(0.0545)	(0.0543)	(0.0554)	(0.0609)	(0.0674)

Standard errors in parentheses

+ p¡0.1, * p¡0.05, **p¡0.01, *** p¡0.001

Table 9: Falsification test regressions with outcomes prior to secondary school.

	Maths		BMI	
	(1)	(2)	(3)	(4)
	Age 7	Age 16	Age 7	Age 16
Reduced-form				
$1[A_i \geq c]$	0.0172	0.0439**	-0.2020	-0.1564
	(0.0181)	(0.0158)	(0.1735)	(0.2505)
Tot. obs.	2731	2707	2597	2259
Obs. in bdwith	1955	1503	1644	1654
Bandwidth	1.848	1.321	1.547	1.852
Without bias correction				
Grammar	0.0376	0.1000**	-0.4522	-0.3363
	(0.0392)	(0.0344)	(0.3882)	(0.5368)
Tot. obs.	2731	2707	2597	2259
Obs. in bdwith	1955	1503	1644	1654
Bandwidth	1.848	1.321	1.547	1.852
First-stage	0.4577***	0.4387***	0.4468***	0.4651***
First-stage s.e.	(0.0290)	(0.0366)	(0.0337)	(0.0319)
With bias correction				
Grammar	0.0491	0.1242**	-0.4329	-0.6094
	(0.0501)	(0.0410)	(0.4792)	(0.7146)
Tot. obs.	2731	2707	2597	2259
Bandwidth	1.2469	1.0388	0.7521	1.2791
Obs. in bdwith	1419	1213	869	1241
First-stage	0.4429***	0.4219***	0.4539***	0.4546***
First-stage s.e.	0.0390	0.0428	0.0493	0.0430

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Summary statistics of proposed mechanisms by type of secondary school attended. Source: NCDS wave 3.

	Grammar		**Sec. modern**	
	Mean	s.d.	Mean	s.d.
More than 50% school children				
– have father in non-manual occupation	0.68	0.47	0.15	0.35
– stayed after school-leaving age pre-1972	0.95	0.22	0.46	0.50
Higher than median				
– % girls taking GCE only (full qualification)	0.95	0.22	0.34	0.4
– % girls taking CSE only (no qualification)	0.31	0.46	0.76	0.43
School lacks				
– library	0.24	0.43	0.23	0.42
– science labs	0.23	0.42	0.35	0.48
– sport facilities	0.31	0.46	0.38	0.49
Observations	1160		2288	

Table 11: Mechanism analysis: Two-stage regressions with channel as the main independent variable and bias correction. Estimates are from rdrobust command in Stata(by Calonico et al. 2014, 2017).

	(1) A-levels	(2) Degree	(3) On benefits	(4) Unemployed	(5) Log wage	(6) High SAH	(7) Low malaise	(8) BMI	(9) Chol	(10) Trig
M1 **50%+ pupils with father non-manual occ**										
RD_Estimate	0.7640+	0.4878	0.1009	-0.2855	0.2728	0.4513	-0.6858	-1.9783	-1.2550	-2.4220
	(0.4556)	(0.3293)	(0.2305)	(0.2344)	(0.7127)	(0.4436)	(0.4322)	(5.0464)	(1.6194)	(3.3932)
Left of c	355	248	263	432	167	335	268	278	207	309
Right of c	456	318	313	510	217	405	335	349	256	377
M2 **50%+ pupils stayed after school-leaving age pre-1972**										
RD_Estimate	0.8892	1.3137	0.2948	-0.4897	-1.1666	0.4503	-3.1562	-36.0040	-0.0335	13.9923
	(0.9532)	(1.7199)	(0.6864)	(0.4234)	(3.5562)	(1.2626)	(11.1763)	(101.0646)	(5.7736)	(81.9422)
Left of c	235	259	301	407	237	307	261	156	165	195
Right of c	281	274	293	411	240	308	275	186	182	214
M3 **Higher than median % girls taking GCE only**										
RD_Estimate	0.5025*	0.3884*	-0.0365	-0.2596+	0.2828	0.0714	-0.1885	-1.2041	-0.4677	-1.2134
	(0.2105)	(0.1600)	(0.1023)	(0.1429)	(0.4186)	(0.2600)	(0.1856)	(2.8831)	(0.6967)	(1.1621)
Left of c	337	378	352	438	280	320	373	325	241	338
Right of c	442	458	405	535	339	407	455	407	295	398
M4 **Higher than median % girls taking CSE only**										
RD_Estimate	-0.7563+	-0.6780+	0.0409	0.3354+	-0.4533	-0.1914	0.4504	0.9677	0.4133	1.6003
	(0.4153)	(0.3691)	(0.1146)	(0.2014)	(0.4572)	(0.3738)	(0.2785)	(2.9436)	(0.7854)	(1.5920)
Left of c	467	375	346	436	248	293	269	335	166	309
Right of c	574	455	402	528	312	374	348	425	224	367
M5 **School lacks library**										
RD_Estimate	2.5685	2.1239	0.1532	-1.2535	7.4610	1.1875	-2.5523	9.1823	7.3167	6.9537

	(3.3157)	(3.4439)	(6.6100)	(1.5665)	(39.8151)	(5.3869)	(5.7110)	(34.2012)	(46.5032)	(14.8207)
Left of c	451	316	471	518	239	344	310	393	225	367
Right of c	557	398	527	609	303	426	393	482	284	440
M6 **School lacks science labs**										
RD_Estimate	-8.4372	-25.9331	0.5642	0.8089	-3.0523	-1.0666	0.7411	4.0828	4.9098	3.3556
	(36.9328)	(412.9228)	(2.1588)	(0.7673)	(5.9060)	(1.2554)	(0.9383)	(11.7094)	(11.9544)	(4.1172)
Left of c	455	383	355	587	209	437	445	388	315	375
Right of c	559	470	410	683	268	534	554	477	371	449
M7 **School lacks sport facilities**										
RD_Estimate	8.3668	-1.2955	0.0903	2.0912	-0.7466	-13.8954	0.9012	126.8574	3.7518	15.5323
	(41.9612)	(1.3100)	(0.5800)	(4.7984)	(1.5316)	(501.8111)	(2.1176)	(3457.3717)	(10.9792)	(54.2559)
Left of c	432	515	338	618	384	371	403	309	252	269
Right of c	538	587	393	721	466	454	499	395	314	331

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$

Appendix

Table A1: Human capital outcomes: reduced form regressions with automatically selected bandwidth. Running variable is the distance between cognitive score and threshold, which is LEA-specific.

	(1) A-levels	(2) Degree	(3) Unemployed	(4) On benefits	(5) Log wage
$1[A \geq c_{LEA}]$	0.121***	0.0663*	0.00119	-0.0202	0.0313
	(0.0365)	(0.0312)	(0.0194)	(0.0225)	(0.0757)
Distance	0.0295	-0.00770	0.00520	0.00702	0.0877
	(0.0309)	(0.0259)	(0.0190)	(0.0181)	(0.0863)
Distance $\times 1[A \geq c_{LEA}]$	0.174***	0.164***	-00702	0.00107	0.00206
	(0.0417)	(0.0355)	(0.0259)	(0.0247)	(0.116)
Non-cognitive skills	0.166	0.159	-0.147**	-0.205***	0.341
	(0.100)	(0.0868)	(0.0556)	(0.0618)	(0.232)
Female	-0.0219	-0.0233	-0561	0.0462***	-0.468***
	(0.0190)	(0.0164)	(0.0101)	(0.0118)	(0.0393)
Mother's interest	0.0346***	0.0185*	0.00201	-0.00992	0.0446*
	(0.0104)	(0.00886)	(0.00556)	(0.00643)	(0.0221)
Father's SES	0.0880***	0.0616***	0.00202	-0.0108	0.0882***
	(0.0111)	(0.00970)	(0.00592)	(0.00695)	(0.0231)
Mother smoke preg.	-0.00184	-0.00283	0.00963	-0.00632	-0.0122
	(0.0114)	(0.00972)	(0.00622)	(0.00696)	(0.0237)
Child morbidity	-0.352	-0.438	0.268	-0.0494	-1.118
	(0.282)	(0.244)	(0.150)	(0.177)	(0.585)
Constant	-0.368***	-0.298***	0.124*	0.309***	1.456***
	(0.103)	(0.0892)	(0.0564)	(0.0637)	(0.232)
F	55.537	30.528	1.518	3.750	22.326
Tot obs. available	2505	2352	2116	2816	1553
Obs. in bandwidth	1599	1538	1213	1849	791
Bandwidth	1.565	1.605	1.330	1.658	1.141
Polynomial	1	1	1	1	1

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2: Health outcomes: reduced form regressions with automatically selected bandwidth. Running variable is the distance between cognitive score and threshold, which is LEA-specific.

	(1) High SAH	(2) Low malaise	(3) BMI	(4) Chol	(5) Trig
$1[A \geq c_{LEA}]$	0.0213	-0.0346	-0.269	-0.176	-0.352
	(0.0698)	(0.0502)	(0.686)	(0.168)	(0.217)
Distance	0.0565	0.0334	1.445	0.188	0.712*
	(0.0980)	(0.0561)	(1.037)	(0.249)	(0.332)
Distance $\times 1[A \geq c_{LEA}]$	0.00388	-0.0665	-2.735*	-0.387	-0.991*
	(0.133)	(0.0754)	(1.375)	(0.333)	(0.442)
Non-cognitive skills	0.212	0.0637	-1.432	-0.176	-0.556
	(0.203)	(0.151)	(2.015)	(0.504)	(0.656)
Female	0.0415	-0.117***	-0.868*	-0.993***	-0.973***
	(0.0364)	(0.0260)	(0.357)	(0.0881)	(0.114)
Mother's interest	0.0322	0.00416	-0.279	-0.0275	0.0887
	(0.0206)	(0.0148)	(0.198)	(0.0488)	(0.0633)
Father's SES	0.0677**	0.0106	-0.917***	-0.0981	-0.146*
	(0.0213)	(0.0153)	(0.212)	(0.0526)	(0.0682)
Mother smoke preg.	-0.0151	-0.0240	0520	-0.0741	0.00420
	(0.0219)	(0.0159)	(0.217)	(0.0524)	(0.0677)
Child morbidity	-0.218	0.330	6.785	0.514	0.442
	(0.523)	(0.381)	(5.129)	(1.281)	(1.658)
Constant	0.0794	0.818***	32.56***	5.282***	3.686***
	(0.201)	(0.149)	(2.031)	(0.505)	(0.660)
F	3.580	3.060	4.678	17.263	10.599
Tot obs. available	1869	1865	1786	1498	1500
Obs. in bandwidth	771	974	716	605	592
Bandwidth	0.913	1.156	0.865	0.875	0.859
Polynomial	1	1	1	1	1

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A3: First stage regressions with automatically selected bandwidths for each outcome. Running variable is the distance between cognitive score and threshold, which is LEA-specific.

	(1) A-levels	(2) Degree	(3) Unemployed	(4) On benefits	(5) Log wage	(6) High SAH	(7) Low malaise	(8) BMI	(9) Chol	(10) Trig
$1[A \geq c_{LEA}]$	0.4637***	0.4535***	0.4361***	0.4444***	0.4507***	0.4692***	0.4608***	0.4717***	0.4434***	0.4388***
	(0.0340)	(0.0344)	(0.0402)	(0.0310)	(0.0514)	(0.0550)	(0.0473)	(0.0574)	(0.0617)	(0.0623)
Distance	0.0461	0.0378	0.0181	0.0440	0.0213	0.0190	-0.0063	0.0364	0.0382	0.0353
	(0.0288)	(0.0286)	(0.0393)	(0.0250)	(0.0587)	(0.0773)	(0.0529)	(0.0868)	(0.0912)	(0.0953)
Distance $\times$	0.1458***	0.1690***	0.2241***	0.1637***	0.2128**	0.1441	0.2178**	0.0989	0.1253	0.1467
$1[A \geq c_{LEA}]$	(0.0388)	(0.0391)	(0.0538)	(0.0340)	(0.0787)	(0.1048)	(0.0711)	(0.1151)	(0.1221)	(0.1268)
Non-cognitive skills	0.2714**	0.3231***	0.3710**	0.2973***	0.4167**	0.5884***	0.5720***	0.3672*	0.2375	0.2491
	(0.0931)	(0.0956)	(0.1154)	(0.0853)	(0.1578)	(0.1601)	(0.1427)	(0.1685)	(0.1848)	(0.1882)
Female	0.0371*	0.0228	0.0243	0.0344*	0.0312	0.0494	0.0323	0.0644*	0.0560	0.0619
	(0.0177)	(0.0181)	(0.0210)	(0.0162)	(0.0267)	(0.0287)	(0.0245)	(0.0299)	(0.0323)	(0.0328)
Mother's interest	0.0517***	0.0463***	0.0606***	0.0510***	0.0677***	0.0353*	0.0524***	0.0534**	0.0567**	0.0593**
	(0.0097)	(0.0098)	(0.0115)	(0.0089)	(0.0150)	(0.0163)	(0.0140)	(0.0166)	(0.0179)	(0.0181)
Father's SES	0.0363***	0.0361***	0.0360**	0.0373***	0.0309*	0.0525**	0.0436**	0.0367*	0.0627**	0.0611**
	(0.0103)	(0.0107)	(0.0123)	(0.0096)	(0.0157)	(0.0168)	(0.0145)	(0.0177)	(0.0193)	(0.0196)
Mother smoke preg.	-0.0029	0.0003	0.0098	-0.0036	0.0120	0.0032	0.0094	-0.0079	0.0006	-0.0002
	(0.0106)	(0.0107)	(0.0129)	(0.0096)	(0.0161)	(0.0173)	(0.0150)	(0.0181)	(0.0192)	(0.0194)
Child morbidity	-0.1116	-0.0096	-0.1463	-0.1204	0.0677	-0.1290	-0.2371	-0.0769	-0.0930	-0.0999
	(0.2628)	(0.2685)	(0.3118)	(0.2435)	(0.3974)	(0.4128)	(0.3595)	(0.4291)	(0.4699)	(0.4754)
Constant	-0.3472***	-0.4009***	-0.4924***	-0.3771***	-0.5561***	-0.6931***	-0.6929***	-0.4705**	-0.4481*	-0.4625*
	(0.0962)	(0.0983)	(0.1169)	(0.0878)	(0.1573)	(0.1586)	(0.1401)	(0.1699)	(0.1853)	(0.1892)
Obs. in bandwidth	1599	1538	1213	1849	791	771	974	716	605	592
Total obs. available	2505	2352	2116	2816	1553	1869	1865	1786	1498	1500

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A4: Falsification test regressions with placebo threshold 0.2.

	(1) A-levels	(2) Degree	(3) Unemployed	(4) On benefits	(5) Log wage	(6) High SAH	(7) Low malaise	(8) BMI	(9) Chol	(10) Trig
Grammar	-1.6035	-17.5806	0.4426	2.0670	-4.8326	2.3686	-17.1055	13.2418	0.0366	-2.5714
	(4.3302)	(257.3159)	(1.5104)	(54.9236)	(11.1316)	(7.1225)	(193.0670)	(48.5393)	(8.0044)	(15.3786)
Distance	1.8397	14.7283	-0.3306	-1.7856	4.2230	-2.1994	16.1229	-11.4886	-0.2491	2.8712
	(4.0777)	(212.9774)	(1.1333)	(47.8273)	(9.3575)	(6.1857)	(181.3791)	(46.5198)	(8.0617)	(15.3199)
Distance ×	-1.5142	-9.1050	0.0959	1.0133	-1.4693	1.9408	-15.1432	13.1347	-0.9405	-5.1000
$1[A \geq c_{LEA}]$	(3.9185)	(132.4068)	(0.5045)	(26.4990)	(3.4829)	(3.5390)	(162.1588)	(51.8202)	(8.0257)	(14.0961)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$1[A \geq c_{LEA}]$	-0.0384	-0.0050	-0.0321	-0.0031	-0.0416	-0.0316	-0.0088	0.0380	0.0352	0.0316
	(0.0824)	(0.0739)	(0.0648)	(0.0816)	(0.0928)	(0.0862)	(0.1013)	(0.1045)	(0.1135)	(0.1148)
First-stage F	0.217	0.005	0.246	0.001	0.201	0.135	0.008	0.132	0.096	0.076
Tot obs. available	2505	2352	2816	2116	1553	1869	1865	1786	1498	1500
Obs. in bdwith	541	604	736	517	390	448	390	387	318	316
Bandwidth	0.489	0.581	0.621	0.542	0.547	0.526	0.459	0.468	0.452	0.444
Polynomial	1	1	1	1	1	1	1	1	1	1

Standard errors in parentheses. Covariates included and omitted from the table.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A5: Human capital outcomes: polynomial regressions for whole sample with donut exclusion around the threshold and automatically selected bandwidth. Running variable is the distance between cognitive score and threshold, which is LEA-specific.

	(1)	(2)	(3)	(4)	(5)
	A-levels	Degree	Unemployed	On benefits	Log wage
Grammar	0.2974**	0.2432*	0.0351	-0.1916**	0.1918
	(0.0922)	(0.1056)	(0.0574)	(0.0734)	(0.1578)
Distance	0.0176	-0.0862	-0.0003	0.1691***	0.0165
	(0.0477)	(0.0775)	(0.0417)	(0.0486)	(0.0744)
Distance $\times 1[A \geq c_{LEA}]$	0.1060	0.1753	0.0091	-0.1245*	-0.0191
	(0.0569)	(0.0922)	(0.0521)	(0.0582)	(0.0940)
Non-cognitive skills	0.0610	0.0705	-0.2571**	-0.1808	0.3232
	(0.1196)	(0.1415)	(0.0800)	(0.0924)	(0.2273)
Female	-0.0236	-0.0383	-0.0104	0.0648***	-0.5135***
	(0.0217)	(0.0242)	(0.0137)	(0.0165)	(0.0380)
Mother's interest	0.0320*	0.0094	-0.0031	-0.0011	0.0241
	(0.0128)	(0.0134)	(0.0079)	(0.0094)	(0.0234)
Father's SES	0.0809***	0.0703***	-0.0041	-0.0172	0.0820***
	(0.0126)	(0.0141)	(0.0079)	(0.0097)	(0.0231)
Mother smoke preg.	0.0096	-0.0047	0.0151	-0.0036	-0.0048
	(0.0129)	(0.0136)	(0.0078)	(0.0094)	(0.0225)
Child morbidity	-0.2817	-0.4086	0.4351*	0.1742	-0.7939
	(0.3186)	(0.3418)	(0.1935)	(0.2339)	(0.5901)
Constant	-0.3029**	-0.2747*	0.2277**	0.3540***	1.4678***
	(0.1152)	(0.1302)	(0.0749)	(0.0869)	(0.2222)
First-stage F	129.757	59.859	60.953	75.366	103.100
Tot obs. available	1250	792	732	999	886
Obs. in bdwith	2393	2246	2020	2691	1481
Bandwidth	1.289	0.846	0.851	0.910	1.409
Polynomial	1	1	1	1	1

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A6: Health outcomes: polynomial regressions for whole sample with donut exclusion around the threshold and automatically selected bandwidth. Running variable is the distance between cognitive score and threshold, which is LEA-specific.

	(1) High SAH	(2) Low malaise	(3) BMI	(4) Chol	(5) Trig
Grammar	0.1735	0.1768	-2.2192	-0.3144	-0.3446
	(0.1611)	(0.1397)	(1.6632)	(0.3672)	(0.4262)
Distance	0.0928	-0.0321	2.5743*	-0.0647	0.4420
	(0.1041)	(0.0995)	(1.3131)	(0.2668)	(0.3476)
Distance $\times 1[A \geq c_{LEA}]$	-0.2323	-0.1062	-3.0800	0.2474	-0.6214
	(0.1289)	(0.1203)	(1.5782)	(0.3393)	(0.4354)
Non-cognitive skills	0.2131	-0.0501	-0.8653	0.1910	-0.0055
	(0.2295)	(0.2064)	(2.4500)	(0.5638)	(0.6497)
Female	0.0119	-0.1309***	-0.9441*	-1.0320***	-0.9113***
	(0.0369)	(0.0325)	(0.4010)	(0.0922)	(0.1078)
Mother's interest	0.0203	0.0074	-0.0305	-0.0087	0.0400
	(0.0218)	(0.0187)	(0.2303)	(0.0557)	(0.0638)
Father's SES	0.0659**	0.0053	-0.8991***	-0.0810	-0.0785
	(0.0225)	(0.0196)	(0.2417)	(0.0592)	(0.0682)
Mother smoke preg.	-0.0108	-0.0175	0.0012	-0.0681	0.0054
	(0.0219)	(0.0190)	(0.2325)	(0.0530)	(0.0608)
Child morbidity	0.4578	0.1470	7.4267	0.1741	0.1344
	(0.5492)	(0.4716)	(5.7194)	(1.3623)	(1.5788)
Constant	0.0815	0.8404***	32.2234***	4.7845***	2.8581***
	(0.2130)	(0.1886)	(2.3322)	(0.5528)	(0.6334)
First-stage F	63.115	56.644	52.136	57.308	49.406
Tot obs. available	751	699	616	551	494
Obs. in bdwith	1778	1774	1696	1421	1422
Bandwidth	0.975	0.900	0.838	0.921	0.822
Polynomial	1	1	1	1	1

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$