Article:

Pastore, Chiara and Jones, Andrew Michael orcid.org/0000-0003-4114-1785 (2023) Human capital consequences of missing out on a grammar school education. Economic Modelling. 106414. ISSN 0264-9993

https://doi.org/10.1016/j.econmod.2023.106414

Economic Modelling 126 (2023) 106414

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling

Human capital consequences of missing out on a grammar school education[☆]

Chiara Pastore[a,b], Andrew M. Jones[a,*]

[a] Department of Economics and Related Studies, University of York, 5 Alcuin Way, Heslington, YO10 5DD York, United Kingdom
[b] Boston Consulting Group, Calle de Alcala 95, 28009 Madrid, Spain

ARTICLE INFO

JEL classification:
I1
I26
I28
C21
C26

Keywords:
Selective schooling
Human Capital
Health
Instrumental Variables
Regression Discontinuity Design

ABSTRACT

This paper disentangles the effect of selective schooling on long-term human capital from that of individual background. The study design proxies entry test scores for selective secondary schools in England with historical data, estimating discontinuities in school assignment directly from the data. We find that, for the marginal admitted student, selective school attendance positively affects educational attainment. Adult labour market and health outcomes are not affected for the marginal admitted student. The effect on educational outcomes is conditional on having a favourable background, and partly explained by higher-ability peers and single-sex schools.

1. Introduction

In 2018, the UK government announced the first £50 million round of a £200 million fund for an expansion of existing grammar schools, public and selective high-quality institutions. Proponents of tracking policies, that allocate students to different classes or schools on the basis of ability, maintain that they reward talent regardless of socio-economic background. Opponents, on the other hand, are concerned that selectivity is skewed in favour of children from affluent backgrounds and children whose ability develops earlier, since they are disproportionately likely to do well in entry tests.

Being assigned to a selective school could affect long-term outcomes through higher peer ability, or through a curriculum with more academic content, which could facilitate later admission to better higher education. Moreover, there is evidence that more qualified teachers seek schools with higher ability pupils and that better resources are allocated to these schools (Pop-Eleches and Urquiola, 2013). Recognising its relevance to the current policy context, this paper explores the medium- and long-term effect of going to grammar school, compared to its main alternative within a selective system, on a broad range of human capital and health outcomes for individuals of similar prior ability. The analysis is based on data from the National Child Development Study (NCDS), a British cohort study of individuals born in March 1958, who started secondary school in 1969 and whose lives have been followed for over 60 years.

The literature looking at the effects of selective schooling can be divided into two main strands. A first set of studies compares selective and non-selective systems, generally finding no difference in average outcomes, but instead a link between selection and inequality in education and earnings (Atkinson et al., 2006; Burgess et al., 2020). The second set of studies on the effect of selective education has estimated a local average treatment effect for the marginal admitted student, based on regression discontinuity design approaches (Clark, 2010; Del Bono and Clark, 2016; Guyon et al., 2012). Due to data limitations, this second approach has not been implemented before with English country-wide data.

☆ We thank Nigel Rice, Thomas Cornelissen, Matt Sutton, Sandra McNally and Emma Tominey, the editor, guest editors and two anonymous reviewers for helpful suggestions and comments, as well as participants to the 2019 RES Conference at the University of Warwick, the HERU Seminar at the University of Aberdeen, the Essen Health Conference, the XXVIII AEDE Meeting at the University of Las Palmas de Gran Canaria and the 2022 Empirical Health Economics Workshop, Ispra. Chiara Pastore acknowledges support by the Economic and Social Research Council (grant number ES/J500215/1) and the comments received at the 2019 EuHEA Student-Supervisor Conference at the Católica Business School. The employers, funders, data creators and UK Data Service have no responsibility for the contents of this paper. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

* Correspondence to: Department of Economics and Related Studies, Heslington, York YO10 5DD, United Kingdom.
E-mail address: andrew.jones@york.ac.uk (A.M. Jones).

https://doi.org/10.1016/j.econmod.2023.106414
Received 20 November 2022; Received in revised form 13 June 2023; Accepted 20 June 2023
Available online 28 June 2023

We use an instrumental variable (IV) strategy that is inspired by a regression discontinuity design (RDD), based on the fact that admission to grammar school was determined by a local pass mark for the entry exam, known as the 11-plus.[1] 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. To overcome the issue of limited data on entry score cut-offs for all English LEAs in the 1960s, we proxy pass marks with LEA-specific thresholds estimated directly from the NCDS dataset. Following the structural breaks literature (Bai, 1997), in the same spirit as Card et al. (2008), we select the threshold value that maximises the fit of a model of school assignment. This approach identifies the medium-term effect on educational attainment and labour market outcomes, as well as the long-term effect on health and risk of developing illness up to age 50.

This paper contributes to the literature on the long-term effect of grammar schools in England and Wales in several ways. Our approach helps us identify the treatment effect for the specific group of pupils who would have been affected by the expansion in grammar school places, if it had happened back then. Second, we work in a context with limited information, constructing the assignment variables and score cut-offs from large survey data, in the absence of administrative records. We use several robustness checks to increase confidence in the study design, confirming previous results obtained with other methodologies. 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. Lastly, since schools are surveyed as part of the NCDS study, we can explore how specific school features may explain any effect of school type for long-term outcomes.

For the marginal student in each area (an area corresponding to a Local Education Authority), we find a significant and positive effect of grammar attendance on the probability of achieving A-levels, a secondary academic qualification, and having a university degree. However, this effect is conditional on having high socio-economic status or high parental interest in education. No effect is found on adult labour market outcomes, health and biomarkers for risk of developing chronic illnesses. This result holds both when using local polynomial regressions and when using bias-corrected procedures proposed by Calonico et al. (2014b, 2017). We also find that peer quality and single-sex education could be significant mechanisms to attain better educational qualifications. Our work draws from historical data and concludes that pupils who narrowly missed out on grammar places in the 1960s missed out on marginally higher chances of better educational attainment, but only if they had a favourable family background.

The paper is structured as follows. We describe the context and existing literature in Section 2, while the data are presented in Section 3. Section 4 discusses the threshold estimation procedure, accompanied by the necessary conditions for identification and the empirical approach. Section 5 presents the results and robustness checks, followed by an exploratory mechanism analysis in Section 6. Section 7 discusses our findings and Section 8 concludes.

2. Background

2.1. Historical context

Grammar schools have been present in England since the 16th Century, but their formal inclusion in the compulsory state-funded secondary education system dates back to the 1944 Education Act. The Act established a tripartite system of secondary schooling, distinguishing between selective grammar schools, non-selective secondary modern schools and a minority of technical schools. The main distinguishing features, apart from the higher average ability of pupils in grammar schools, were the level of academic content of the curriculum and the different educational and occupational prospects for pupils. Since the Act, admission to grammar schools has been determined by performance in the 11-plus exam, testing language, numerical and reasoning skills. Until the 1960s, the vast majority of children would sit the 11-plus at the end of primary school, usually in September of their last year. The test was set locally by Local Education Authorities (LEAs), and the entry mark depended on the number of grammar school places in the area. Rather than setting a given pass score, LEAs usually considered the distribution of scores and then assigned grammar school places to children scoring in the top tail, so the pass mark varied every year. On average, pupils scoring in the top 25% of the distribution in their local area were admitted to grammar school (Bolton, 2017). Area-specific differences in admission included different school capacity constraints, and different policies concerning teacher's recommendations, distance from the school and having other siblings already at the school. Panels of teachers and LEA representatives made decisions on where to allocate students. Those who did not reach the entry score were generally assigned to secondary modern schools.[2] Parents could subsequently request a different school or decide to appeal against the panel's decision if they did not agree with it, but this was uncommon. Thus, students with similar ability scores could be assigned to different types of school for two reasons: either because they were close to the cut-off entry score for their area, or because they were from different areas. To the extent that ability slightly above or below the entry cut-off is random, this paper can isolate the long-term effect of going to grammar school, compared to just missing out on admission, on a broad range of outcomes. Identification therefore focuses on the marginal pupil within each LEA.

2.2. Related literature

Regression discontinuity designs (RDD) based on assignment test scores have 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; Dustmann et al., 2017; Guyon et al., 2012; Kirabo Jackson, 2010; Pop-Eleches and Urquiola, 2013). 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. In another study, Guyon et al. (2012) exploit a reform causing an exogenous increase in the number of pupils attending grammar schools in Northern Ireland, and find that average educational outcomes increased following the reform. A third study, by Del Bono and Clark (2016), is the one most similar to ours, since they investigate long-term effects of elite schools in Scotland, for children born in the 1950s. In addition to educational and labour market outcomes, they study effects on fertility. They find a significant and positive effect on educational outcomes, while they observe higher wages and lower fertility for girls only. Their RDD strategy exploits information on admission exam scores (equivalent to the 11-plus), and entry score cut-offs for the city of Aberdeen.

[1] As both the assignment variable and cutoff are proxied, standard RDD inference and treatment concepts may not apply. We therefore adopt an instrumental variable approach for inference and definition of the local average treatment effect (LATE).

[2] From 1965 onwards the Circular 10/65 approved by the Labour government encouraged LEAs to move towards a third type, comprehensive schools, catering for students of all abilities. This paper focuses on areas that were still largely selective when the data was collected, and where a large proportion of pupils attending public secondary schools were assigned to either grammar or secondary modern. See Section 5.5 for further discussion.

Given the unavailability of such precise information at Local Education Authority level for England, literature on the effects of grammar schools has been unable to exploit RDD methodologies to answer this question in the English context, except for the study by Clark (2010), which focuses on one region in England and on education outcomes only. The main challenge in the literature has been to deal with selection bias, arising from the inability to control exhaustively for pre-treatment characteristics of treated and control children. Value-added and instrumental variable approaches, such as Galindo-Rueda and Vignoles (2005) and Harmon and Walker (2000), have been criticised by Manning and Pischke (2006) as unable to remove the selection problem, which was evidenced by a placebo-type test, showing a spurious effect of type of school on outcomes prior to school.

More recent studies focusing on the English context have tried to overcome this problem by matching treated and control groups on either individual or local-area characteristics. Among these, Atkinson et al. (2006) find a positive effect of grammar on educational outcomes, while Burgess et al. (2020) find that selection increases educational and earnings inequality. Fewer papers have focused on the health and well-being effects of selective education. Although the raw data shows better adult health for former grammar pupils, the effect of grammar on health is mostly not statistically significant when including all relevant pre-treatment characteristics (Jones et al., 2012). However, the average could be hiding heterogeneity by ability levels. For instance, the transition to comprehensive schooling has been found to worsen health and smoking for individuals with lower non-cognitive skills only (Basu et al., 2018). We now build further on this evidence by looking at human capital for the group of marginal students affected by the additional places. The effect is policy relevant as it expresses the impact of making it into grammar school versus missing out on the additional marginal place within each LEA, for similar levels of initial ability and everything else being equal.

3. 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 nationally representative for that cohort. Following the birth sweep, which contained over 17,000 individuals, surveys were undertaken at ages 7, 11, 16, 23, 33, 42, 45, 50, 55. At the latest sweep the survey still retained over 9000 individuals. For the present study, only English and Welsh individuals were included in the sample, since Scotland had a different schooling system in place.[3] A discussion of the possible implications of attrition in the data is in Appendix for the interested reader.

3.1. Sample and type of school

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. Information on LEA of school, essential for recreating the actual peer group of test-takers from NCDS data, was obtained via special licence access, also from the age 16 wave.[4] Of the total 18,521 individuals for whom there is information in the survey, 5366 were excluded because they did not have relevant information on school attended, while a further 9131 individuals were excluded as they did not attend neither grammar nor secondary modern school leaving 4024 individuals. This large drop in the sample is due to the fact that many areas at the time were already transitioning to a mixed-ability system following the Circular 10/65, which encouraged LEAs to abolish selection by ability at the school level. 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 3448 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 1450 and 2800, depending on the outcome. The reason why we do not use the same sample for threshold estimation and treatment effect estimation is that a reliable threshold figure needs to capture as much of the actual test-taking population as possible. Once the location of the threshold is estimated, we estimate the first and second stage of the 2SLS on the same sample, see details in Section 4.

3.2. Ability scores

The main variable needed for the identification strategy, age 11 cognitive ability score, 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 et al., 2011). More details on this procedure can be found in the Appendix. The cognitive ability index is the first principal component obtained via PCA, which explains approximately 85% of the total variance in the three test scores. This approach is preferred to including the three separate scores to rule out multicollinearity issues in the estimation procedure, on the one hand. On the other, the three scores obtain similar relative weights for the first principal component obtained via PCA, meaning their contribution to our index is roughly equal. The 11-plus actual total score was also an average of similar language, mathematics and reasoning tests, offering further support to the assumption that our index is a good proxy of 11-plus scores.[5] 97% of NCDS children took the NCDS ability tests between April and July of 1969, thus only a few months after having sat the 11-plus (see Appendix Figure A1). This is usually taken in September of the last year of primary school, which for NCDS children would have been 1968–1969. For our empirical application, it is reassuring that ability tests were still taken in primary school, and not at the start of secondary school, where tests could be affected by the treatment.

3.3. Outcomes

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. We measure these outcomes on different waves of the NCDS to capture completed formal education up to university degree level (at age 23), labour market outcomes at a point where both graduates and non-graduates are established in the labour market (at age 33), and health outcomes in mid-life (age 45–50), when gradients in chronic conditions have begun to emerge. Education outcomes, collected at age 23, are binary variables equal to 1 for having obtained any A-levels (or equivalent) and having a university degree, and 0 otherwise. Labour market outcomes, measured at age 33, are retrieved from survey questions and include binary variables for being unemployed, receiving state benefits (excluding child benefits) and gross hourly wage, which is calculated from weekly hours worked and weekly wages and then log-transformed for regression analysis. Self-reported health outcomes include age 50 self-assessed health and low malaise scores, which have been validated as good predictors of

[3] Just under 3% of the individuals in the working sample are Welsh.

[4] To understand how many individuals may cross the border to go to school in another LEA, we compare age 16 LEA of school with age 7 LEA of residence. Only 8% of individuals in our sample are recorded in different LEAs.

[5] Similarity of the two tests is also confirmed by comparing the individual tasks. For the interested reader, we suggest comparing the sweep 2 NCDS ability test, available on the Centre for Longitudinal Studies website, with sample tests in Bristow (2016), containing a range of tests used from the 1950s onwards.

physical and mental health. Self-assessed health (SAH) is measured on a 5-point scale ranging between 'Excellent' and 'Very poor', and converted to a binary variable equal to 1 if SAH is 'Excellent' of 'Very good', and 0 otherwise. Malaise score is measured via the 9-item Malaise Inventory (see Appendix). In the regression analysis, we use a binary variable equal to 1 for low malaise (score 0–2 out of 9) and 0 otherwise, increasing in good mental health. We also include the following three biomarkers measured at age 45, all increasing in the risk of cardiovascular disease and health complications. Body mass index (BMI) is calculated as weight in kg divided by squared height in meters, while cholesterol ratio (mmo/L) and triglyceride levels (mmo/L) are retrieved from blood samples.

3.4. Background information

Our identification strategy relies on a discontinuity with respect to cognitive ability but we do control for other covariates within the regression models. These controls are also used to check the robustness of the findings and to explore heterogeneity in the local treatment effects. The vector of relevant covariates includes sex; mother's interest in child education on a 4-point scale and father's socio-economic status on a 5-point scale; whether the mother was smoking during 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, indicating no morbidity, and 1, indicating highest morbidity (Jones et al., 2011). We further 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).[6] We then converted the total BSAG score to a variable bounded between 0 and 1 for convenient interpretation.

4. Methods

4.1. Threshold selection

The threshold for each local education authority, necessary for identification, is estimated directly from the data. If the data were available, with pupils' individual test scores (at the 11-plus exam) and the actual LEA cutoffs that were used to determine access to grammar schools, we could implement a standard Regression Discontinuity Design (RDD). Without these data: (i) we must use a proxy variable for the true test scores using the ability tests provided in the NCDS data; (ii) we do not know the exact cutoff in each LEA and we have to estimate these thresholds. There are studies that have addressed the properties of RDD estimators when there is measurement error in the running variable and when the cutoff is unknown, but we are not aware of studies that deal with both problems together. For example, Porter and Yu (2015) discuss inference when the unknown cutoff is estimated using the true (not a proxy) running variable and a key finding is that the RDD treatment effect can be efficiently estimated with discontinuity points estimated from the data. While, Dong and Kolesár (2023) address the issue of measurement error by redefining the treatment effect so that it relates to the observed proxy rather than the true running variable. They show that, provided the observed running variable classifies treatment assignment correctly and affects the conditional means of the potential outcomes smoothly, then ignoring the measurement error yields the local average treatment effect where the observed running variable is equal to the cutoff. They suggest a donut-hole approach to broaden the robustness of this approach.

Our approach is to use an instrumental variable (IV) strategy. This is motivated by a fuzzy RDD and estimation is done in the usual way by two stage least squares (2SLS). However, as we use a proxy running variable and estimated cutoffs, we draw on an IV setting to interpret our estimates as local average treatment effects (LATE). We exploit a discontinuity in the effect of the NCDS ability score (used as a proxy for the true running variable) on the probability of entering grammar school. This discontinuity is estimated from the data. The motivation and intuition for this that the ability proxy could be considered as mapping the true running variable into another ability measure. As in a standard RDD approach, we control for a flexible polynomial in the proxy of ability both in the first and the second stages (as ability is obviously associated with better human capital outcomes) to preserve the exclusion restriction assumption, while the estimated discontinuity in the proxy variable is used as the exclusion restriction. In short, this non-linearity should only predict grammar school entry, and not other pre-grammar school individual characteristics, conditional on a polynomial in the proxy ability score. Once controlling for a polynomial in test scores in the first and second stage, in this IV interpretation, the compliers will be the marginal grammar school entrants.

Thus, for each LEA, we first run probit models for grammar attendance with a single regressor $1[A_i \geq c]$. This is an indicator function for whether individual i's ability A_i is equal or greater than 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 et al. (2008), who look for the presence of tipping points (i.e. discontinuities) in changes in the share of white population in US neighbourhoods over time, as a function of share of other minorities. Their hypothesis is that discontinuities in these changes are located at specific values of the base-year minority population share. The study analyses a larger sample than ours, and 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.[7] Compared to Card et al. (2008), we have the advantage that an LEA-specific discontinuity is known to be present in the school assignment function (since there was a pass mark for grammar school entry), although its exact location is not observed.

4.2. Identification

Identification of treatment effects is based on the assumption that pupils scoring near the LEA-specific cutoff, where the discontinuity is observed, have similar baseline characteristics. If individuals on both sides are similar enough, 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 cut-off.

Identification is based on two stages. The first stage models school assignment. We denote the treatment variable as $G_i \in \{0, 1\}$, where

[6] The twelve attributes are measures of social maladjustment and include unforthcomingness, withdrawal, depression, anxiety for acceptance by adults, hostility towards adults, 'writing off' of adults and adult standards, anxiety for acceptance by children, hostility towards children, restlessness, 'inconsequential' behaviour, miscellaneous symptoms and miscellaneous nervous symptoms.

[7] Alongside the 'structural break' approach we use, Card 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.

$G_i = 1$ indicates grammar attendance. Following Lee and Lemieux (2010), treatment assignment, which is assumed to change discontinuously at a LEA-specific cut-off 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 the indicator variable for equal or greater than the threshold, $h(.)$ is a generic function of individual's distance from the pass mark and υ a random error term.[8] We expect this function to be non-deterministic, since 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 11-plus test scores by themselves did not grant access to grammar school, but other factors contributed to admission too, as discussed in Section 2.[9] Second, additional fuzziness may be caused by limitations of the data, as noted by Card and Giuliano (2016) in their analysis of 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, although we provide evidence in support of the threshold selection procedure performed, we acknowledge that not observing actual threshold values may further increase fuzziness in the first stage. We therefore rely on a fuzzy RDD, where the probability of treatment assignment does not jump sharply from 0 to 1 at the threshold, but by a smaller amount.

The second-stage equation 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 Eqs. (1) and (2). Following previous literature, we use an estimator analogous to a Wald estimator in 2SLS procedures (Hahn et al., 2001; Lee and Lemieux, 2010), so that the local average treatment effect (LATE) 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 a local 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. 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 for 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 goodness of fit measured on the available data, and we would expect this to be reflected in the density function of the constructed assignment variable (see Appendix Figure A2). 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. Figures A3 and A4 in the Appendix show that this assumption holds in two different checks.

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 and smoothness of covariates as a function of the assignment variable), β in Eq. (2) can formally be interpreted as a local average treatment effect (LATE) for individuals in the proximity of the measured threshold (Lee and Lemieux, 2010). 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 cut-off 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 an expansion in grammar school places back then, as students on the margin would have been those most likely affected by it. At the same time, under the stated assumptions, the estimation strategy allows us to isolate the historical effect of grammar attendance from pre-schooling ability and other pre-treatment confounders.

4.3. Implementation

4.3.1. Choice of bandwidth, kernel and polynomial

A way to ensure similarity between treated and control group is to accurately choose the neighbourhood around the cut-off, from which observations for the estimation of β are drawn. Following the standard RDD literature, we refer to this neighbourhood as the bandwidth h (Lee and Lemieux, 2010). The smaller the bandwidth, 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 cut-off will yield estimates with lower variance but potentially higher bias (Lee and Lemieux, 2010).[10] 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, $MSE = (\hat{\beta} - \beta)^2$ (Imbens and Kalyanaraman, 2012; Calonico et al., 2014b).

While an 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 et al., 2014b, 2017, 2018, 2020; Cattaneo and Vazquez-Bare, 2016). The argument in a nutshell is that in using the estimated MSE-optimal bandwidth for

[8] Our approach follows a standard "normalising and pooling" transformation, where the assignment scores are centred around the cutoff for each LEA. Fort et al. (2022) consider a normalising and pooling design where the treatment is delivered at different "sites" and a fixed number of places are allocated from the highest score until they are exhausted. This creates a problem with an endogenous cutoff, where there is one observation located exactly at each site-specific threshold and, consequently, a violation of the continuity of the pooled density of the assignment variable at the zero-normalised cutoff. Our application does not fall into this trap and we do not observe a violation of continuity in the density of the pooled running variable. This is because we do not use the actual test scores and instead rely on proxies. Also, the site (LEA) specific cutoffs are generated by a grid search over fixed values rather than being based on the actual test scores of our respondents. It is also worth stressing that the NCDS cohort respondents (limited to those born in first week of March 1958) are only a small fraction of the full year group that were ranked and assigned to grammar schools within each LEA. As a result, our application falls under what Fort et al. (2022) define as a "non-problematic multi-cutoff design" with the cutoff exogenously set. We note that the solution proposed by Fort et al. (2022), when the problem of endogenous cutoffs does arise, is to include site fixed effects and that this is the "safest and likely efficient option" in that case. In our analysis we cluster the standard errors at LEA level.

[9] Similar factors are also described in Del Bono and Clark (2016) and Clark (2010), although they both rely on assignment scores only to predict admission.

[10] 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.

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, a procedure known as undersmoothing (McCrary, 2008; Cattaneo and Vazquez-Bare, 2016). Instead, Calonico et al. (2014b) propose a robust bias-correction procedure (CCT correction from here onwards) for bandwidth selection, which estimates bias and then adjusts both the regression discontinuity (RD) point estimate and variance estimator. The CCT correction allows for robust confidence intervals, less sensitive to small bandwidth variations and accounting for the variability introduced when correcting for the estimated bias term in the treatment effect estimator.[11] Alongside our main specification, with local regressions implemented using the MSE-optimal bandwidth selection approach, we thus propose an alternative with CCT bias correction, in order to investigate sensitivity of the analysis to this procedure.[12]

The empirical implementation further requires kernel functions and polynomials 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 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 cut-off 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, 2019), we refrain from using higher orders, and only introduce an interaction term between threshold and distance from threshold in the main specification. This accounts for the intuition that not only the intercept but also the slope of the average of the dependent variable as a function of the assignment variable may be different above the threshold.[13]

4.3.2. Empirical specification

The empirical counterparts to Eqs. (1) and (2) are 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 distance and the discontinuity and X_i is a vector of individual characteristics. We estimate Eqs. (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 et al. (2014a).

Finally, we run the analysis adding an interaction term between the treatment indicator and sex, high father's SES and high mother interest in child education, in order to explore heterogeneity in the effect of grammar school. We prefer this to subsample analysis, given the small sample sizes in our study. We also note that each interaction term will be endogenous. For the model to be identified, we thus predict the interaction $Grammar \times Female$ with an interaction between the original discontinuity indicator and sex, expressed as $T \times Female$, and so on for the other characteristics.

5. Results

5.1. Summary statistics

Summary statistics for baseline characteristics and outcomes by type of school for the whole sample are shown in Table 1. Grammar pupils display higher average cognitive and non-cognitive 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 considered. They have a 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 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 the treatment indicator (grammar) as a single regressor first, and then with additional covariates, estimated on the whole available sample (see Table A1 in the Appendix). The coefficient for grammar is highly significant for all outcomes, although, in line with expectations, some of this association is accounted for by adding covariates to the model. The only exception is the malaise indicator, not significantly linked with grammar. For all outcomes, the association is explained mainly by cognitive skills and the five included covariates. When we estimate the LATE via the RDD approach, focusing on the neighbourhood close to the threshold, we aim at netting out the effect of these pre-treatment characteristics. As discussed above, in proximity of the threshold, these do not vary discontinuously, while the probability of treatment does. Reassuringly, when we contrast characteristics by type of school for the whole sample to those of compliers, we find that grammar and secondary modern pupils are more similar among the latter (see Table A2 in the Appendix, where the complier group is defined as the sample of observations within MSE-optimal bandwidths for A-level outcome).

5.2. Discontinuity in grammar school assignment

Fig. 1 displays LEA-specific threshold values selected by the threshold search procedure outlined in Section 4.1. Higher estimated thresholds are not necessarily found in areas with the least grammar places. Instead, the location of the threshold reflects the observed LEA-specific

[11] Details on the CCT correction procedure and the theory behind it, including the bias and standard error estimators, can be found in Calonico et al. (2014b), with a more recent exposition in Calonico et al. (2020). An alternative approach to MSE-optimal bandwidth, introduced by Calonico et al. (2020), aims at minimising 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 et al. (2020) 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.

[12] In a recent study, Hyytinen et al. (2018) exploit a feature of the Finnish seat assignment mechanism in local elections to compare standard and CCT-corrected RD estimates to experimental estimates: they find that CCT correction produces closer estimates to the experimental ones, while the standard procedure based on MSE-minimising bandwidth yields biased results. While this result is to some extent specific to that context, research practice in RDD applications is moving towards adopting CCT correction procedures.

[13] For formal details on the choice of bandwidth, kernel and polynomial and the squared error minimisation procedure with the inclusion of covariates see Calonico et al. (2017, 2018).

Table 1
Descriptive statistics by 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	1	4	1.88	1.03	1	4
Father's SES	3.32	0.90	1	5	2.82	0.81	1	5
Mother smoke pregnancy	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	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
Biomarkers (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			

Mother interest in child education is on a scale from 1-Little interest to 4-Over concerned. Father's SES is on a scale from 1-Low to 5-High. Maternal smoking during pregnancy is on a scale from 1-Non-smoker to 4-Heavy smoker. Healthy ranges for the biomarkers are: <25 for BMI, <5 for cholesterol ratio, <1.7 for triglycerides.

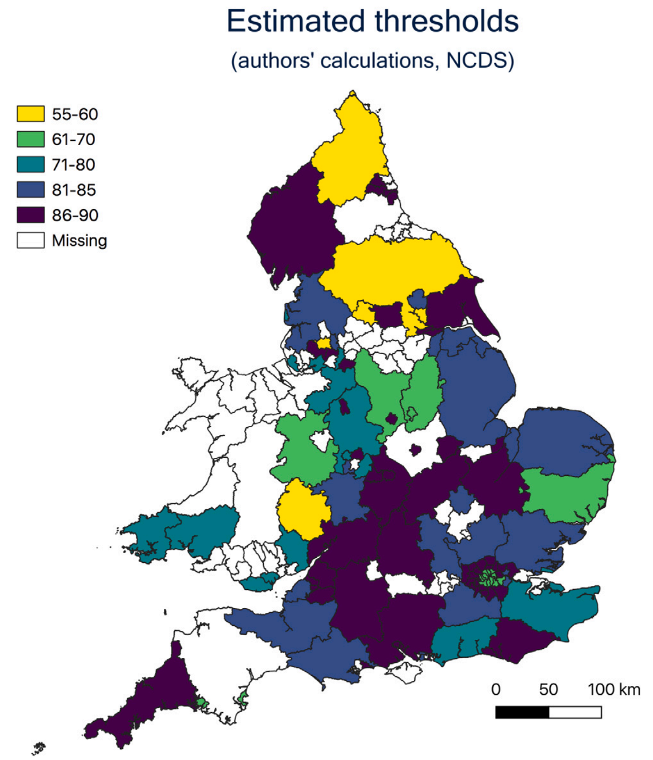

Fig. 1. Threshold values estimated through our "structural breaks" approach, for the English and Welsh LEAs in the data. The values are expressed in terms of the percentile of the NCDS cognitive ability distribution that would give access to a grammar school place in 1969.

discontinuity, as inferred from the ability of grammar pupils from each LEA in our sample. The figure shows that in our historical sample, the discontinuity in grammar admission was found at lower values of the assignment scores for LEAs in the North, with a couple of exceptions. The discontinuity was found at highest values of the assignment scores in the Midlands, in the South and in Cornwall. A traditional grammar school stronghold such as Kent displays a discontinuity at a lower value, possibly reflecting the larger number of grammar school places in this region. Fig. 2 plots grammar attendance in pre-specified bins, for different levels of the distance between the assignment variable and the LEA-specific cut-off, roughly expressing the probability of grammar attendance as a function of the distance variable. The figure shows a jump in the probability of treatment when the assignment variable is near the threshold, indicated by the dashed vertical line, where $A_i - c_{LEA} = 0$. 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. The instrument of interest, the indicator variable $1[A \geq c_{LEA}]$, is highly predictive of the first stage and sizeable. Estimating the first stage using each of the 10 outcomes in turn, including all covariates, we find a coefficient on the instrument 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 (full results in Table A3 in the Appendix). 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. Yet, as long as these covariates are smooth around the threshold, this association does not threaten identification of the treatment effect of interest.

5.3. Effect of grammar schools for the marginal student

Fig. 3 shows scatter graphs for each outcome, so that each dot represents the average outcome for groups of individuals at similar levels of the distance variable. While all outcomes display a relationship with the assignment variable in the expected direction, none shows a

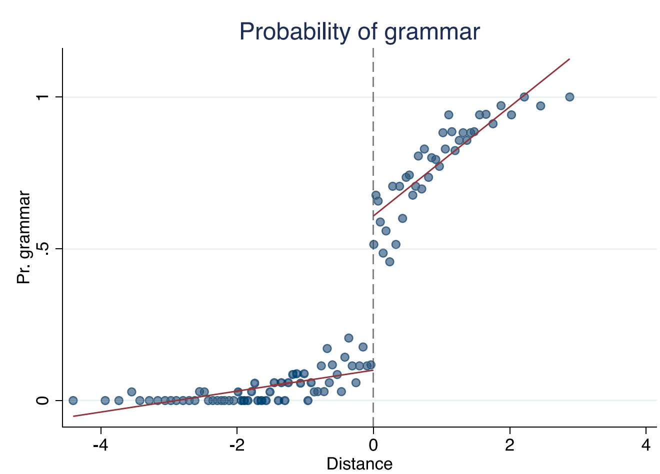

Fig. 2. Scatter graph for probability of grammar school attendance as a function of the distance D_i between the assignment variable A_i and the LEA-specific threshold c_{LEA}. Observations are grouped in 50 bins, yielding average probability of treatment within the group. Average bin size is 69 (N = 3448).

sharp jump at the threshold value of zero. The probability of achieving A-levels and a degree display a fairly similar pattern, increasing steeply and linearly after the threshold. Most other outcomes only show an average improvement after the threshold, more evident for cholesterol ratio, triglycerides and the indicator of excellent or very good SAH, but no evident discontinuity.

Table 2 shows LATE estimates for education, labour market and health outcomes, obtained via estimation of Eqs. (3) and (4), for the sample within MSE-optimal bandwidths, without CCT bias correction.

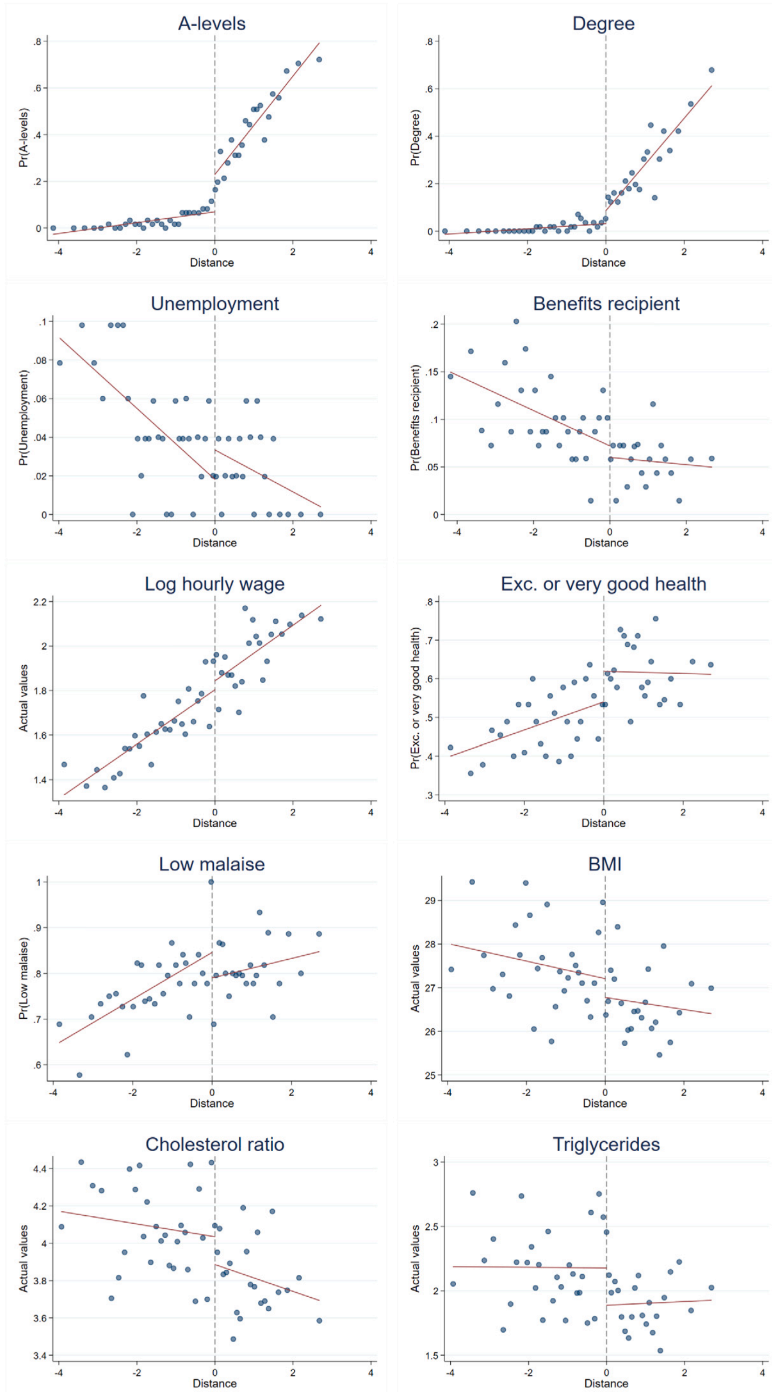

Fig. 3. Outcomes as a function of distance between the assignment variable and the LEA-specific threshold, $D_i = A_i - c_{LEA}$. Observations are grouped in 50 bins, yielding average outcome within the group.

All models include the five covariates described in Section 3. 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 (Lee and Lemieux, 2010). 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 ($p < 0.05$). For both A-levels and obtaining a university degree, the interaction variable between distance and threshold indicator is positive and significant. This indicates a positive change in the slope of the probability of both outcomes as a function of the ability score after the threshold, already shown in Fig. 3. This result is consistent with the fact that grammar schools were the track with highest academic focus.

The estimated LATE of grammar school for adult labour market, health and disease risk outcomes is not significant, despite being mostly of the expected sign. The sign of grammar is negative for the probability of receiving benefits, cholesterol ratio and triglycerides, while it is positive for hourly wages and self-assessed health. The grammar coefficient for unemployment is close to 0, while the one associated to the probability of scoring low malaise is negative. The interaction

Table 2
Human capital and health outcomes: local polynomial regressions with pre-selected bandwidth.

	(1) A-levels	(2) Degree	(3) Unemployed	(4) On benefits	(5) Log wage
Grammar	0.2607***	0.1461*	0.0027	−0.0454	0.0693
	(0.0753)	(0.0678)	(0.0442)	(0.0504)	(0.1665)
Distance	0.0175	−0.0132	0.0052	0.0090	0.0863
	(0.0319)	(0.0273)	(0.0194)	(0.0195)	(0.0879)
Distance × $1[A \geq c_{LEA}]$	0.1357***	0.1398***	−0.0007	0.0085	−0.0127
	(0.0402)	(0.0360)	(0.0270)	(0.0253)	(0.1166)
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 statistic	185.989	174.073	117.957	205.661	76.830
Obs. in bandwidth	1599	1538	1213	1849	791
Total obs. available	2505	2352	2116	2816	1553
	(6) High SAH	(7) Low malaise	(8) BMI	(9) Chol	(10) Trig
Grammar	0.0454	−0.0750	−0.5705	−0.3980	−0.8025
	(0.1479)	(0.1085)	(1.4425)	(0.3814)	(0.5000)
Distance	0.0556	0.0329	1.4656	0.2031	0.7406*
	(0.0993)	(0.0555)	(1.0634)	(0.2598)	(0.3472)
Distance ×$1[A \geq c_{LEA}]$	−0.0027	−0.0502	−2.6789*	−0.3369	−0.8735*
	(0.1320)	(0.0763)	(1.3536)	(0.3331)	(0.4437)
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 statistic	72.710	95.030	67.538	51.711	50.402
Obs. in bandwidth	771	974	716	605	592
Total obs. available	1869	1865	1786	1498	1500

Standard errors clustered at LEA level in parentheses. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

between distance and the threshold indicator is significant for BMI and triglycerides level.

Results when implementing the CCT bias-correction procedure are fairly consistent with those estimated without bias correction (see Appendix Table A4).[14] The CCT bias-corrected MSE-optimal bandwidths are generally smaller, except for hourly wage, cholesterol ratio and triglyceride levels. With the bias-correction procedure, grammar attendance increases the probability of obtaining A-levels by 25 percentage points ($p < 0.01$) and the probability of obtaining a university degree by 19 percentage points ($p < 0.05$). As before, the grammar coefficient is not significant for health outcomes and risk of illness. Discontinuity estimates in the grammar assignment function (first-stage estimates), at the bottom of each panel, are all highly significant with bias correction also, offering support to the validity of the study design.

5.4. Heterogeneous effects

In Tables 3 and 4, we additionally provide results for the specification with treatment interacted with sex, father's SES and mother's interest in child education. Now the grammar coefficient expresses the effect of grammar attendance for the base category in each regression. In panel A of Table 3, the grammar coefficient on A-levels is larger than with the pooled sample, and the coefficient on the interaction $Grammar \times Female$ is negative and significant. Thus, for boys, the base category, attending grammar increases the probability of achieving any A-levels by 30 percentage points, while the effect is 11 points lower for girls. Additionally, the coefficients in column (2) indicate that the effect of grammar on the probability of achieving a university degree is only significant for boys, the base category. In panels B and C, grammar attendance only has a positive effect on probability of achieving A-levels for children whose father has high SES (18 percentage points increase) and whose mother's interest in child education is high (12 percentage points increase). Conversely, the effect of grammar attendance for the base categories, children with low SES or low mother interest, is not significant, indicating that grammar school attendance only gives an educational advantage to advantaged groups. Again, results for health do not show a significant impact of grammar school, except for cholesterol ratio, which is significantly positively impacted for children with high mother interest. Finally, a brief note on how these observed characteristics affect grammar school admission (i.e. our first stage) can be found in the Appendix.

5.5. Robustness checks

We include several robustness checks of our approach. We first implement local polynomial regressions for a placebo cut-off, in order to show that the first stage is not predictive of grammar attendance at other points of the distribution of the assignment variable (see Figure A5 and Tables A5 and A6).

As a second robustness check, we re-estimate the model while excluding a set of observations around the threshold, a procedure known as donut exclusion or donut-RDD. The density of the distance from LEA-specific threshold $D_i = A_i - c_{LEA}$ presents a concentration of observations around the threshold, and we show that our results are not sensitive to excluding these (see Appendix Figure A1 and Appendix Tables A7 and A8).

Thirdly, we estimate the effect of the discontinuity on placebo outcomes prior to secondary school. This check is in the same spirit as an influential test conducted by Manning and Pischke (2006). The authors show that a set of empirical approaches to estimate the effect of grammar school did not deal successfully with selection bias, since outcomes prior to secondary school appeared to be affected by secondary school attendance. Our approach deals successfully with this difficulty and the placebo procedure is illustrated in detail in the Appendix.

A final point on robustness concerns the changing landscape of secondary schools in the 1960s and 1970s. When NCDS pupils went to secondary school, LEAs had started to transition to a mixed-ability system, establishing 'comprehensive' secondary schools that catered

[14] The CCT bias-corrected estimates are obtained with the 'rdrobust' package for Stata (Calonico et al., 2014a, 2017).

Table 3
Human capital outcomes: local polynomial regressions with pre-selected bandwidth and treatment interacted with individual characteristics.

	(1) A-levels	(2) Degree	(3) Unemployed	(4) On benefits	(5) Log wage
Panel A: by sex					
Grammar	0.3193***	0.1574*	−0.0056	−0.0541	0.0670
	(0.0804)	(0.0723)	(0.0476)	(0.0535)	(0.1770)
Female	0.0191	−0.0172	−0.0072	0.0404*	−0.4727***
	(0.0298)	(0.0259)	(0.0165)	(0.0188)	(0.0663)
Grammar × Female	−0.1128*	−0.0220	0.0155	0.0169	0.0046
	(0.0522)	(0.0472)	(0.0306)	(0.0338)	(0.1225)
First-stage F	92.950	86.906	58.976	102.736	38.254
Panel B: by father's SES					
Grammar	0.1918*	0.1231	−0.0061	−0.0419	0.0621
	(0.0789)	(0.0717)	(0.0481)	(0.0537)	(0.1866)
High father's SES	0.0368**	0.0409**	−0.0030	−0.0069	0.0821**
	(0.0142)	(0.0130)	(0.0080)	(0.0093)	(0.0308)
Grammar × High father SES	0.1877***	0.0701	0.0229	−0.0097	0.0183
	(0.0435)	(0.0408)	(0.0262)	(0.0291)	(0.1036)
First-stage F	89.612	84.768	55.139	98.660	34.628
Panel C: by mother interest					
Grammar	0.1884*	0.1366	0.0245	−0.0452	0.1403
	(0.0953)	(0.0859)	(0.0557)	(0.0611)	(0.2017)
High mother's interest	0.0022	0.0083	−0.0028	−0.0146	0.0584
	(0.0189)	(0.0168)	(0.0110)	(0.0119)	(0.0445)
Grammar ×	0.1184*	0.0303	−0.0230	−0.0217	−0.0567
High mother interest	(0.0536)	(0.0491)	(0.0321)	(0.0345)	(0.1211)
First-stage F	84.403	78.244	55.225	95.514	36.613
Obs. in bandwidth	1599	1538	1213	1849	791
Total obs. available	2505	2352	2116	2816	1553

Standard errors clustered at LEA level in parentheses. High father SES corresponds to mid-high or highest. High mother interest corresponds to very interested or over-concerned. Distance, interaction with distance and covariates included and omitted from the table. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4
Health outcomes: local polynomial regressions with pre-selected bandwidth and treatment interacted with individual characteristics.

	(6) High SAH	(7) Low malaise	(8) BMI	(9) Chol	(10) Trig
Panel A: by sex					
Grammar	0.0724	−0.0320	0.0207	−0.1796	−0.7307
	(0.1630)	(0.1170)	(1.6126)	(0.4257)	(0.5580)
Female	0.0592	−0.0798	−0.3935	−0.8175***	−0.8712***
	(0.0617)	(0.0442)	(0.6080)	(0.1498)	(0.1967)
Grammar × Female	−0.0488	−0.0807	−1.0787	−0.3910	−0.1323
	(0.1207)	(0.0822)	(1.2025)	(0.3054)	(0.3996)
First-stage F	36.299	47.387	33.777	25.830	24.787
Panel B: by father's SES					
Grammar	0.0007	−0.0854	−0.7216	−0.3846	−0.8435
	(0.1710)	(0.1227)	(1.6525)	(0.4296)	(0.5585)
Father's SES	0.0440	0.0086	−0.9774***	−0.0658	−0.1207
	(0.0293)	(0.0210)	(0.2940)	(0.0724)	(0.0947)
Grammar × Father SES	0.1006	0.0242	0.3826	−0.0350	0.1098
	(0.1037)	(0.0706)	(1.0382)	(0.2566)	(0.3308)
First-stage F	31.531	42.240	29.909	23.522	22.797
Panel C: by mother interest					
Grammar	0.1114	−0.1013	−1.2000	−0.8292	−1.1177
	(0.1856)	(0.1347)	(1.9207)	(0.4696)	(0.6303)
Mother's interest	0.0223	−0.0169	−0.6520	−0.2041*	0.2061
	(0.0381)	(0.0271)	(0.3840)	(0.0948)	(0.1268)
Grammar × Mother int.	−0.1146	0.0672	0.8221	0.7432*	0.2219
	(0.1159)	(0.0822)	(1.2098)	(0.2925)	(0.3963)
First-stage F	36.130	47.481	31.227	25.112	24.559
Obs. in bandwidth	771	974	716	605	592
Total obs. available	1869	1865	1786	1498	1500

Standard errors clustered at LEA level in parentheses. High father SES corresponds to mid-high or highest. High mother interest corresponds to very interested or over-concerned. Distance, interaction with distance and covariates included and omitted from the table. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

for all abilities.[15] Our analysis focuses on pupils from largely selective areas: the average percentage of LEA-level comprehensive pupils for the LEAs included in our sample is 25% against 41% of the total NCDS sample, illustrating still a high degree of selectivity in our sample, in spite of the reform. Including in our main analysis an indicator for degree of LEA selectivity, obtained from the 1971 edition of the Comprehensive School Committee Journal, does not affect our main findings (results available on request).

6. Mechanism analysis

Finally, we attempt to unpack treatment effect by testing whether specific attributes of grammar schools lead to more favourable outcomes for grammar pupils. We focus on peers' ability and characteristics, teacher quality and school resources, all collected in the school questionnaire of NCDS wave 3. Peer ability is measured by looking at percentages of pupils taking either only GCEs (General Certificate of Education, equivalent to A-levels, of higher academic value) or only CSEs (Certificate of Secondary Education, of lower academic value, not requiring completion of a full standard qualification).[16] To define peer environment, we also take an indicator for whether the school was single sex. For teaching quality, we construct a variable indicating whether teachers get any training for career guidance, and the percentage of teachers who left the school in the previous year. As a proxy of school resources, we define a binary indicator for whether the school lacks any facilities (including library, science labs, sports facilities and other).

The exercise is mainly exploratory in nature, since we do not provide a breakdown of the components of the full treatment effect. Instead, we provide a reduced-form evaluation of the extent to which each mechanism might explain the long-term effect of grammar school attendance. We do this by reproducing the main RDD analysis with a twist: we substitute treatment in the first stage with each of the mechanisms of interest in turn. Since each given school attribute is highly correlated to school type, we expect the threshold for grammar attendance, T_i, to also be predictive of such attribute. We adopt the following two-stage empirical specification:

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

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

In the first stage, the discontinuity used to predict grammar attendance in our main specification (see Section 4.3.2) is now used to predict each proposed mechanism M_i. In the second stage, we isolate the effect of each mechanism on the outcomes. Given that, by construction, each M_i predicted by Eq. (5) is highly correlated with the grammar indicator G_i, μ_1 can be read as an indirect assessment of the proportion of the effect of grammar on the outcomes that is explained by M_i, for the population of compliers.

Summary characteristics for proposed mechanisms by school type are shown in Table 5. In order to implement the RDD procedure as above, we dichotomise all variables before the analysis. We transform the variables for school peers' ability and teachers leaving into binary variables taking value 1 if above the median for the relevant variable, 0 otherwise. 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.[17] 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 the lower CSE qualification. A significantly higher proportion of grammar schools are single sex. Grammar schools display a slightly higher chance that teachers receive training in career guidance for their students, a lower probability of being above the median for proportion of teachers leaving the school and a lower chance that the school lacks facilities.

Table 6 shows results for the mechanism analysis, with MSE-optimal bandwidths. Given that we found a significant effect of grammar on education outcomes only, we focus on these for the mechanism analysis.[18] Peer ability is positively linked to both educational outcomes considered. Attending a school with a proportion of GCE-takers above the median increases the probability of achieving any A-levels by 42 percentage points and a degree by 25 percentage points (p < 0.01

Table 5
Descriptive statistics of mechanisms by secondary school attended.
Source: NCDS wave 3.

	Grammar		Sec. modern		Min	Max
	Mean	s.d.	Mean	s.d.		
Peer environment						
Above median % girls taking GCE only	0.95	0.22	0.34	0.47	0	1
Above median % girls taking CSE only	0.31	0.46	0.76	0.43	0	1
Single sex	0.68	0.47	0.25	0.44	0	1
Teaching and resources						
Teachers get career guidance training	0.85	0.35	0.81	0.39	0	1
Above median % teachers left last year	0.47	0.50	0.52	0.50	0	1
School lacks facilities	0.49	0.50	0.58	0.49	0	1
Observations	1160		2288			

Table 6
Mechanisms: local polynomial regressions with each channel as the treatment variable and pre-selected bandwidth.

	(1) A-levels	(2) Degree
High % girls taking GCE only	0.4172**	0.2478*
	(0.1331)	(0.1101)
First-stage F statistic	41.214	28.683
High % girls taking CSE only	−0.8902*	−0.5080
	(0.4448)	(0.2914)
First-stage F statistic	6.374	5.390
Single sex	0.6837*	0.3487*
	(0.2661)	(0.1673)
First-stage F statistic	11.212	14.759
High % teachers left	−0.9574	−0.6141
	(0.5198)	(0.4043)
First-stage F statistic	5.615	4.482
Teachers get career training	−4.1580	−1.7195
	(6.2892)	(2.2078)
First-stage F statistic	0.457	0.620
School lacks facilities	−3.7742	−1.8606
	(6.6113)	(2.9943)
First-stage F statistic	0.316	0.406
Obs. in bandwidth	1599	1538
Total obs.	2505	2352
Bandwidth	1.565	1.605

Standard errors clustered at LEA level in parentheses.
*$p < 0.05$.
**$p < 0.01$.

[15] The Circular 10/65, promoting this move, was issued four years before NCDS cohort members took the 11-plus, and therefore there were large areas where selection was still in place.

[16] 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. Both the GCE and CSE age 16 examinations were replaced in 1988 with the General Certificate of Secondary Education (GCSE) for pupils of all abilities, while GCE A-levels were kept for ages 16–18 (see webpages at https://qualifications.pearson.com).

[17] 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.

[18] Full results for all outcomes are available on request.

and $p < 0.05$). Being in a school with above median CSE-takers, not requiring completion of the school cycle, is negatively associated with both A-levels and degree, but the association is only significant for A-levels ($p < 0.01$). The single sex feature is also linked to better performance, increasing the chances of A-levels and a university degree by 68 and 35 percentage points respectively ($p < 0.05$). No other mechanism is significantly associated with the two outcomes. However, the first-stage F statistic is below the critical value of 10 in all cases, except for girls taking GCEs and single sex, indicating that the grammar discontinuity is not predictive enough of the other proposed mechanisms. The exploratory mechanism analysis is still informative, as it points towards an effect of some measures of peer ability and single-sex education on education outcomes. Unfortunately, little extra information was available on the quality of teaching and academic content of school curriculum, which remain important channels to consider.

7. Discussion

Results indicate that pupils who miss out on grammar school places within the selective system mainly miss out on higher chances of attaining better educational qualifications. We find a significant effect of grammar school attendance on A-levels and obtaining a university degree across all specifications. The heterogeneity analysis indicates that grammar is likely to only improve educational outcomes for pupils of high SES, or whose education parents are highly interested in. Moreover, the effect on education is larger for boys compared to girls, a finding that corroborates previous evidence (Del Bono and Clark, 2016). In our exploratory mechanism analysis, we linked the effect of grammar school attendance to peer environment. Having a high percentage of school peers taking a high-level qualification and being in a single-sex institution, predicted using the discontinuity for grammar school admission, increase the probability of achieving A-levels and a university degree. Since the magnitude of the coefficient for grammar is smaller than those for the two mechanisms (see Tables 2 and 6), we may even say that, for our group of compliers, these mechanisms explain more of the variation in outcomes than school type alone. For instance, the importance of the single-sex feature has been widely documented in the returns to education literature, and it could arguably also play an important role in this context (Sullivan et al., 2011).

When isolating its effect from confounders, grammar attendance is not a significant predictor of labour market outcomes, health or risk of developing illness for the marginal admitted student. Our evidence corroborates other RDD studies that find significant and positive effects of higher quality schools on educational outcomes within a selective system, while not necessarily for labour market or health outcomes (Dustmann et al., 2017; Clark, 2010; Del Bono and Clark, 2016). Instead, other literature has shown that health and health behaviours may be significantly affected by school duration and attainment, with one of the latest papers on this topic published by the present journal (Baltagi et al., 2023). Thus, raw differences observed in Table 1 must be explained by background characteristics other than school, since they become insignificant around the threshold, where ability and other background factors are more homogeneous. In most cases, the sign of the insignificant grammar coefficient is as expected, and Fig. 3 suggests that significance of the grammar indicator is likely to increase as we move away from the threshold, where confounders would bias the estimate. We additionally note that OLS results are surprisingly similar to those obtained via the RDD approach (see Appendix Table A1). This may suggest that controlling for observed characteristics already isolates the effect of grammar to a reasonably good extent, or that, in our sample, heterogeneity in observables is relatively low.

The null result may be somewhat surprising for wages, given that grammar is expected to increase the probability of achieving higher educational qualifications, which should translate into better wages. A possible explanation for this is that grammar pupils scoring just above the threshold do not rake up the benefits from higher educational qualifications, being of lower ability than the average grammar pupil. For example, they could end up in lower-ranked universities, or do worse in their degrees. If this is true, their wages may end up similar to those for pupils scoring just below the threshold who attended secondary modern, especially given that the latter provided a more practical education that could offer a streamlined way into specific jobs (e.g. well-paid blue-collar occupations, administrative positions). An alternative explanation for the insignificant coefficient could be that the sample is too small to yield enough precision in the coefficient estimate.

In support of the first interpretation, we cite another study using NCDS by Brunello and Rocco (2017). The authors find that for pupils with lower qualifications (approximately 11–12 years of schooling), wages are initially higher for individuals with vocational qualifications compared to those with academic qualifications, and that this pattern is reversed by age 50. Linking this to our findings, it might be that pupils who make it into grammar and achieve lower qualifications experience a wage disadvantage to start with, ending up with similar wages to secondary modern pupils at age 33, as the wage profiles start to overlap. For higher qualifications (approximately 14–15 years of schooling), Brunello and Rocco (2017) find no significant differences in wage profiles by education type over the life-course, which also aligns with our results.

The first £50 million round of the Selective Schools Expansion Fund announced in 2018 created 2700 new grammar school places in 16 schools for the following academic year (UK Department for Education, 2019). With 163 grammar schools present in England at the time of writing, this means that approximately 1 in every 10 grammar schools was allowed to expand. In judging the potential of this programme to reward talent rather than background, two key considerations to be made concern who can access grammar schools and what impact they can make. Previous literature has answered the first question, showing that pupils from privileged backgrounds are up to 45 percentage points more likely to attend grammar than pupils from deprived ones (Burgess et al., 2018). Addressing the second issue, our findings show that past generations of grammar pupils on the margin of being admitted have benefitted from grammar school attendance only in terms of a higher probability of attaining A-levels and possibly higher chances of a university degree. Moreover, we have shown that this effect was likely to apply only to pupils of high socio-economic status or whose family was highly supportive of their education. For the group of compliers, other measures of long-term human capital and health are to be linked to other background factors. Our findings are based on the lives of individuals who faced very different circumstances from current generations of pupils. For instance, obtaining A-levels and a university degree is now more common than it used to be, and grammar schools are not the only public institutions offering a more academic education. However, this historical evidence may still be relevant, since the expansion policy affects the intake and resources of both selective and non-selective schools in selective areas, potentially bringing back some aspects of the old system. On another note, due to the nature of the identification strategy, we did not explore the implications of the policy for pupils who are very far below the threshold, which may shed light on other important consequences of an expansion in selective school places.

8. Conclusion

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 to estimate its location 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 in the 1960s did not benefit in terms of long-term human capital accumulation, with the exception of the direct positive effect on education outcomes, which are conditional on having a favourable background. A more prominent role in explaining raw differences by type of school might be played by the child's cognitive skills and background.

Our research shines a light on the topic of grammar schools in England, finding no evidence that this system can by itself significantly improve human capital regardless of socio-economic background. In fact, when comparing groups of pupils with similar ability at age 11 (both groups on the margin of being admitted), attending a grammar school only affected the likelihood of obtaining A-levels and a University degree, and only for pupils of higher socio-economic background. We acknowledge that further research with data from more recent cohorts is needed to assess the impact of this policy on new generations of pupils, who today face a very different world of education and work, and a broader general knowledge about positive health behaviours. Further research could additionally 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.

Declaration of competing interest

We declare that we have no competing interests.

Data availability

The authors do not have permission to share data.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.econmod.2023.106414.

References

Abdulkadiroglu, A., Angrist, J.D., Pathak, P.A., 2014. The elite illusion: Achievement effects at boston and new york exam schools. Econometrica 82 (1), 137–196.
Atkinson, A., Gregg, P., McConnell, B., 2006. The Result of 11 Plus Selection: An Investigation into Opportunities and Outcomes for Pupils in Selective LEAs. CMPO Working Paper No. 06/150, University of Bristol.
Bai, J., 1997. Estimation of a change point in multiple regression models. Rev. Econ. Stat. 79 (4), 551–563.
Baltagi, B.H., Flores-Lagunes, A., Karatas, H.M., 2023. The effect of higher education on women's obesity and smoking: Evidence from college openings in Turkey. Econ. Model. 123.
Barreca, A.I., Lindo, J.M., Waddell, G.R., 2016. Heaping-induced bias in regression-discontinuity designs. Econ. Inq. 54 (1), 268–293.
Basu, A., Jones, A.M., Rosa Dias, P., 2018. Heterogeneity in the impact of type of schooling on adult health and lifestyle. J. Health Econ. 57, 1–14.
Bolton, P., 2017. Grammar School Statistics. House of Commons Briefing Paper No. 1398, House of Commons Library.
Bristow, M., 2016. the Eleven-Plus Book: Genuine Exam Questions from Yesteryear, third ed. Michael O'Mara Books.
Brunello, G., Rocco, L., 2017. The labor market effects of academic and vocational education over the life cycle: Evidence based on a british cohort. J. Hum. Cap. 11 (1), 106–166.
Burgess, S., Crawford, C., Macmillan, L., 2018. Access to grammar schools by socio-economic status. Environ. Plan. A: Econ. Space 50 (7), 1381–1385.
Burgess, S., Dickson, M., Macmillan, L., 2020. Do selective schooling systems increase inequality? Oxford Econ. Pap. 72 (1), 1–24.
Calonico, S., Cattaneo, M.D., Farrell, M.H., 2020. Optimal bandwidth choice for robust bias corrected inference in regression discontinuity designs. Econom. J. 23 (2), 192–210.
Calonico, S., Cattaneo, M.D., Farrell, M.H., Titiunik, R., 2017. Rdrobust: Software for regression-discontinuity designs. Stata J. 17 (2), 372–404.
Calonico, S., Cattaneo, M.D., Farrell, M.H., Titiunik, R., 2018. Regression discontinuity designs using covariates. Rev. Econ. Stat. 101 (3), 442–451.
Calonico, S., Cattaneo, M.D., Titiunik, R., 2014a. Robust data-driven inference in the regression-discontinuity design. Stata J. 14 (4), 909–946.
Calonico, S., Cattaneo, M.D., Titiunik, R., 2014b. Robust nonparametric confidence intervals for regression-discontinuity designs. Econometrica 82 (6), 2295–2326.
Card, D., Giuliano, L., 2016. Can tracking raise the test scores of high-ability minority students? Amer. Econ. Rev. 106 (10), 2783–2816.
Card, D., Mas, A., Rothstein, J., 2008. Tipping and the dynamics of segregation. Q. J. Econ. 123 (1), 177–218.
Cattaneo, M.D., Vazquez-Bare, G., 2016. The choice of neighborhood in regression discontinuity designs. Obs. Stud. 2, 134–146.
Cawley, J., Conneely, K., Heckman, J., Vytlacil, E., 1997. Cognitive ability, wages, and meritocracy. In: Devlin, B., Fienberg, S.E., Resnick, D.P., Roeder, K. (Eds.), Intelligence, Genes, and Success: Scientists Respond To the Bell Curve. Springer New York, New York, NY, pp. 179–192, Chap. 8.
Clark, D., 2010. Selective schools and academic achievement. B.E. J. Econ. Anal. Policy 10 (1), 1–38.
Del Bono, E., Clark, D., 2016. The long-run effects of attending an elite school: Evidence from the UK. Am. Econ. J.: Appl. Econ. 8 (1), 150–176.
Dobbie, W., Fryer, R.G., 2014. The impact of attending a school with high-achieving peers: Evidence from the new york city exam schools. Am. Econ. J.: Appl. Econ. 6 (3), 58–75.
Dong, Y., Kolesár, M., 2023. When can we ignore measurement error in the running variable? J. Appl. Econometrics.
Dustmann, C., Puhani, P.A., Schönberg, U., 2017. The long-term effects of early track choice. Econom. J. 127 (603), 1348–1380.
Fort, M., Ichino, A., Rettore, E., Zanella, G., 2022. Multi-cutoff RD Designs with Observations Located at Each Cutoff: Problems and Solutions. IZA DP No.15051, IZA Institute of Labor Economics.
Galindo-Rueda, F., Vignoles, A., 2005. The Heterogeneous Effect of Selection in Secondary Schools: Understanding the Changing Role of Ability. Centre for the Economics of Education Working Paper No. 52, London School of Economics..
Gelman, A., Imbens, G., 2019. Why high-order polynomials should not be used in regression discontinuity designs. J. Bus. Econom. Statist. 37 (3), 447–456.
Guyon, N., Maurin, E., McNally, S., 2012. The effect of tracking students by ability into different schools: A natural experiment. J. Hum. Resour. 47 (3), 684–721.
Hahn, J., Todd, P.E., Van der Klaauw, W., 2001. Notes and comments: Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica 69 (1), 201–209.
Harmon, C., Walker, I., 2000. The returns to the quantity and quality of education: Evidence for men in England and Wales. Economica 67 (265), 19–35.
Hyytinen, A., Meriläinen, J., Saarimaa, T., Toivanen, O., Tukiainen, J., 2018. When does regression discontinuity design work? Evidence from random election outcomes. Quant. Econ. 9 (2), 1019–1051.
Imbens, G., Kalyanaraman, K., 2012. Optimal bandwidth choice for the regression discontinuity estimator. Rev. Econom. Stud. 79 (3), 933–959.
Jones, A.M., Rice, N., Rosa Dias, P., 2011. Long-term effects of school quality on health and lifestyle : Evidence from comprehensive schooling reforms in England. J. Hum. Cap. 5 (3), 342–376.
Jones, A.M., Rice, N., Rosa Dias, P., 2012. Quality of schooling and inequality of opportunity in health. Empir. Econ. 42 (2), 369–394.
Kautz, T., Heckman, J.J., Diris, R., Weel, B.T., Borghans, L., 2014. Fostering and Measuring Skills : Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success. OECD Education Working Papers No. 110, pp. 1–87.
Kirabo Jackson, C., 2010. Do students benefit from attending better schools? Evidence from rule-based student assignments in trinidad and tobago. Econom. J. 120 (549), 1399–1429.
Lee, D.S., Lemieux, T., 2010. Regression discontinuity designs in economics. J. Econ. Lit. 48 (2), 281–355.
Manning, A., Pischke, J.S., 2006. Comprehensive versus selective schooling in England in Wales: What do we know? NBER Working Paper No. 12176.
McCrary, J., 2008. Manipulation of the running variable in the regression discontinuity design: A density test. J. Econometrics 142 (2), 698–714.
Pop-Eleches, C., Urquiola, M., 2013. Going to a better school: Effects and behavioral responses. Amer. Econ. Rev. 103 (4), 1289–1324.
Porter, J., Yu, P., 2015. Regression discontinuity designs with unknown discontinuity points: Testing and estimation. J. Econometrics 189 (1), 132–147.
Sullivan, A., Joshi, H., Leonard, D., 2011. Single-sex schooling and labour market outcomes. Oxford Rev. Educ. 37 (3), 311–332.
UK Department for Education, 2019. Selective schools expansion fund: successful applications, 2018 to 2019. Transparency data. [online] Accessed 25th November 2019. URL https://www.gov.uk/government/publications/selective-schools-expansion-fund-successful-applications-2018-to-2019.