Does selection by ability in secondary schooling affect future health and well-being of those who do not make the cut?

Andrew M. Jones*†, Chiara Pastore*, Nigel Rice*‡

May 2018

Abstract

This paper explores the effect of selection by ability in secondary schooling at age 11 on adult health and well-being, adding a timely piece of evidence to the current policy debate on the expansion of selective schools in England. We estimate two separate treatment effects: that of going to a high-ability selective school (grammar), and that of going to a low-ability school (secondary modern), compared to attending a mixed-ability school (comprehensive). The data, from the National Child Development Study cohort, is first pre-processed through matching, and then analysed via parametric regressions, in order to address selection bias. Our models are estimated via both ordinary least squares (OLS) and instrumental variable (IV) approaches, although the former is preferred as we do not find robust evidence against exogeneity of school type. Outcomes include adult self-assessed physical and mental health; BMI; biomarkers for risk of cardiovascular disease; life aspirations in adolescence; adult life satisfaction, self-worth, crime and drug use. Attendance to grammar or secondary modern school does not affect most of our adult health and well-being outcomes, when compared to attending comprehensive. Childhood cognitive and non-cognitive abilities, conversely, are significant and positive predictors of health and well-being in adulthood.

Keywords Health, Selective schooling, Educational Reform, Matching, Instrumental variables.

JEL I1, I26, I28, C21, C26.

*Department of Economics and Related Studies, University of York, UK

†Centre for Health Economics, Monash University, Victoria, Australia

‡Centre for Health Economics, University of York, UK

[0]Corresponding author: Chiara Pastore. Email:cp798@york.ac.uk. Address:A/D/246 Department of Economics and Related Studies, York YO10 5DD, United Kingdom. We thank participants at the HEDG seminars at the University of York, the NCDS 60 Years of Our Lives conference, and the RES Junior Symposium 2018 at the University of Sussex for helpful comments.

1 Introduction

Prevention strategies to reduce the incidence of health problems are among the key priorities for governments and health systems alike, given the pressure represented by the demand for care by ever-increasing and ageing populations. The role of education as a determinant of demand for health and healthcare was one of the key contributions of Michael Grossman's 1972 seminal paper in health economics. In his model, education is thought to increase the individual's valuation of health benefits on the one hand, and their health productivity on the other, thus raising their optimal level of investment in health (Grossman, 1972). Based on the idea that the earlier the investment in human capital, the larger the returns over one's lifetime, implementing sound educational policies for the childhood and teenage years can be an effective way to improve health and living standards at older ages (Campbell et al., 2014). If individuals are given the tools and knowledge to be more efficient producers of health, they can take up a more active role in keeping themselves healthy. On the other hand, educational improvements that are not strictly focused on increasing health knowledge might also improve health, by boosting children's cognitive and non-cognitive skills, shaping their preferences over time and risk, determining their future work and therefore resource control, their peer networks and habit formation (Fuchs, 1982, Cutler and Lleras-Muney, 2011). In this sense, it becomes interesting to explore whether policies that shift educational quality, such as introducing selection on the basis of ability in secondary schooling, have spill-over effects on long-term health and well-being.

A large part of the literature on the association between education and health has focused on the empirical estimation of the relationship between years of schooling and health, which is generally positive, although not always significant, and larger when estimated with lower starting levels of education (Albouy and Lequien, 2009; Arendt, 2005; Ayyagari et al., 2011; Oreopoulos, 1997; Silles, 2009). The advantage of exploiting reforms affecting school leaving age is that they constitute an exogenous change in schooling, which allows researchers to make substantiated claims about causality in the education-health link. However, focusing on years of education alone might miss important aspects of the relationship between schooling and health (Quis and Reif, 2017). Furthermore, as the extension of compulsory schooling reaches its maximum enforceable limit, it is paramount to assess the impact of different policy tools to improve education for the population. Our paper fits in the literature addressing these concerns, while aiming at providing timely evidence to inform the current policy debate on the reintroduction of selective schools in England[1].

We assess the long-term impact on health and well-being of passing or not passing selection in secondary school versus attending a non-selective system. The empirical analysis relies on data from the National Child Development Study (NCDS), a British cohort study of individuals born in March 1958, giving a complete picture of cohort members' lives through a broad range of variables collected over time, including high-quality information on their education, health and personal history. The paper exploits the comprehensive schooling reform implemented in England and Wales in the 1960s, a time in which some NCDS children were exposed to a selective system, while others attended school in a comprehensive system, where school assignment did not depend on ability. The system experienced largely depended on Local Education Authority (LEA) of

[1]Note that several other countries incorporate selection in their secondary schooling systems, including the United States, Australia, Germany and the Netherlands.

residence. Attendance to different systems exposed pupils to different curricula, teacher quality and peer ability, thus offering an opportunity to explore the effect of variation in quality. The study has been used in the past to inform the debate on selective schooling in terms of its impact on education and labour outcomes (Kerckhoff, 1986, 1996, Galindo-Rueda and Vignoles, 2005). This literature has been criticised as unable to eliminate selection bias arising from pre-existing differences between treatment and control pupils (Bonhomme and Sauder, 2011; Manning and Pischke, 2006). Using a similar approach to that implemented by Manning and Pischke (2006), we are able to show that our analysis allows us to overcome such problems, increasing confidence in the robustness of our results. We build on the literature exploring health impacts of the comprehensive reform in the 1960s (Basu et al., 2018; Jones, Rice, and Rosa Dias, 2011; Jones, Rice, and Rosa Dias, 2012), updating previous results with new methods and measures. Our study is unique in linking quality of schooling, defined by selection at a young age, and biometric markers for cardiovascular disease and stress, as well as a broader concept of well-being in later adulthood, in a quasi-experimental framework. We find that type of secondary school attended does not affect most of our adult health and well-being outcomes, while childhood cognitive and non-cognitive abilities have a long-lasting impact on several adult outcomes.

Section 2 briefly reviews the history of selective schooling in the English education system, and the existing knowledge on its effects. This is followed by a detailed description of the data used in Section 3. Section 4 outlines a reference framework for the relationship between schooling and later outcomes. We then describe a two-step approach implemented to estimate the treatment effects of interest, overcoming the main challenges in the literature. Matching methods are combined with a parametric model, in order to yield 'doubly robust' estimates (Ho et al., 2007). Section 5 presents the main results, along with appropriate robustness checks. Section 6 discusses these findings and Section 7 concludes.

2 Background

2.1 Selection by ability in England

At the time of writing, 163 grammar schools exist in England, attended by approximately 167,000 pupils (Richardson, 2016). The origin of selection in the British secondary schooling system traces back to the 1944 Education Act, which established the reorganisation of state secondary schools by Local Education Authorities (LEAs) in a tripartite systems. Pupils could access grammar schools, of highest academic quality, conditional on their performance in the 11-plus test, taken in the last year of primary school, usually at age 11, and on the school's capacity constraints. If they did not pass the exam, they would usually attend secondary modern schools, less academically demanding, geared towards trades. The third type, technical schools, mainly for vocational training, did not require an exam and were not particularly prevalent. Grammar schools admitted on average the top 25% of the cognitive ability distribution in their local area (Richardson, 2016). Entry tests consisted of different modules, including numerical and verbal reasoning, English comprehension, punctuation and grammar, non-verbal reasoning and creative writing (Richardson, 2016). Schools could draft their own exams, so difficulty varied across schools. Official Government statistics report that in 1955 about 28% of state secondary school pupils attended grammar, 64% attended secondary modern, 4%

attended technical, and 4% other schools (Bolton, 2016).

Throughout the 1950s however, dissatisfaction started to increase around the allocation system in state schools. Firstly, 11-plus performance did not reflect fairly pupil ability, but rather preparation for a specific test format. This was often achieved by private tutoring, which discriminated according to family socioeconomic status, given its high monetary cost. Secondly, there were egalitarian concerns around the idea itself of separating pupils by ability at such a young age, given the long-lasting inequality arising from this early life sentence to a more or a less privileged future. With Circular 10/65 in 1965, the Labour government started phasing out the selective system of secondary schooling. While lacking compelling power, the Circular strongly encouraged LEAs to present plans to create comprehensive schools that catered for all abilities, or to convert existing grammar to comprehensive. Because of the non-compulsory nature of the Circular, the phase out was gradual and, in general, areas with a Conservative political majority were slower in adopting the comprehensive system (Galindo-Rueda and Vignoles, 2005, Bolton, 2016). In 1998, with the School Standards and Framework Act, the Labour government outlawed establishment of any new schools that selected pupils by ability (Richardson, 2016). The Act also made provisions for local ballots on the future of existing grammar schools, although these were not implemented everywhere.

In 2015, after a four-year long campaign, Weald of Kent Grammar School received approval to expand in Sevenoaks district, marking the first approval for an expansion of a selective school in years. In September 2016, the newly appointed Prime Minister Theresa May pledged support for new selective schools as a means towards a meritocratic education system, thus starting a new era for selective schools in England. The 2017 UK Spring Budget reported a forthcoming £320 million investment for 140 new free schools, including selective, on top of the 500 already promised by 2020 (Richardson, 2016). After a step back on lifting the grammar school ban following the loss of the Conservative majority in the June 2017 elections, a recent announcement concerning a £50 million investment for grammar schools brings them back to the policy agenda(Long et al., 2017). This changing policy environment thus opens space for research seeking evidence on the potential impact of a more selective education system for present and future generations of pupils.

2.2 Long-term effects of selective schooling

The rationale for selection by ability is that resources and curricula can be targeted more efficiently when groups have homogeneous cognitive abilities. In practice however, experience shows that better resources are attracted by schools admitting higher-ability pupils (e.g. cream-skimming of teachers), widening even more the capability and opportunity gap between pupils going to high-ability and lower-ability schools. Children with higher cognitive abilities at age 11, who are admitted to high-tier schools, might benefit from more challenging programmes and better teaching, which might help develop their cognitive and non-cognitive skills more, as well as introducing them to more able and socially connected networks. Selective schools traditionally also place more emphasis on sports and other extra-curricular activities, which are known to help personal development and socio-emotional skills. We start from the premise that higher skills, better work and income opportunities, positive peer effects and better habit formation that come with a higher quality education can all affect health and well-being, in the short and possibly long term. Comparatively, those who do not make the cut are likely to miss out on these resources and on related spill-over effects. Additionally, the psychological effect of missing the entry score for the high-ability school can have repercussions on the child's self-confidence, shaping their ambitions and plans for the future.

The 1960s comprehensive reform in England offers an opportunity to evaluate long-term effects of selection by ability in secondary schooling, a topic of interest for the UK as well as other countries applying similar systems. Yet, the lack of a clear rollout of the reform has made it difficult to isolate its effect on life outcomes from other confounding factors. A standard problem in the returns on education literature is the endogeneity of schooling, often due to unobservable factors that determine both education quality or attainment and other outcomes, or due to simultaneity, both of which result in selection bias in treatment effect estimates. The literature has dealt with this issue in different ways, mainly in the estimation of effects on earning and educational achievement. Using NCDS data, Galindo-Rueda and Vignoles (2005) explore the effect of the comprehensive reform on achievement at school, by instrumenting comprehensive school attendance with political control in the county and share of comprehensive schools in the individual's LEA. Their results suggest that the shift to comprehensive schooling reduced achievement for more able children only. Using a nationally representative British household panel, Burgess et al.(2014) compare selective and non-selective LEAs to investigate the impact of selection on earnings inequality, and find that inequality in average hourly wage is significantly higher in selective areas. Burgess et al. (2017) further investigate access and completion of University education by area and type of school within the area. They find that within selective areas, grammar pupils had significant higher chances of accessing and completing higher education, as well as attending a high-status University. When compared to non-selective area pupils with similar school exam scores however, grammar students did not do significantly better at University.

The validity of this type of analysis was put under scrutiny by Manning and Pischke (2006), who criticise comparisons of outcomes for pupils going to school in selective and comprehensive areas as unable to remove the selection problem arising from fundamental differences between the two groups. Firstly, they argue that a comprehensive located in the same areas as a grammar experiences the same shortcomings characterising secondary modern schools. If this is true, then comprehensive schools also admit pupils with lower average ability than grammar schools, and estimation of treatment effects using comprehensive as control group will be biased[2]. Secondly, Manning and Pischke (2006) do find an effect of comprehensive attendace on post-secondary school maths score: going to comprehensive reduces maths test scores by approximately 0.08 standard deviations. However, they point out that an indicator of comprehensive school attendance is also (unexpectedly) a significant predictor of maths scores at age 11, before secondary schooling, controlling for age 7 achievement. They interpret this as a 'falsification test' showing that the apparently significant effect of selective schooling for post-school scores is in fact due to underlying differences in average characteristics of pupils in selective and comprehensive areas (see appendix). They also note that this result does not change when implementing instrumental variable (IV) strategies using political control of the county as instrument. These results are later endorsed by Bonhomme and Sauder (2011), who find that, when using a difference-in-differences approach to correct for unobservables, the effect of selective schooling on test scores disappears.

[2]A solution could be to include only purely selective and purely comprehensive LEAs in the analysis. However, the same authors argue that average characteristics across these non-mixed LEAs are still too different to adequately identify a causal effect.

A number of studies have used alternative methods that are more robust to the criticisms advanced above. Using Scottish data, Del Bono and Clark (2014) implement a regression discontinuity design (RDD) to estimate the impact of elite schools on educational attainment, income and fertility for the marginal student[3]. By using entry test score cutoffs to model probability of elite school attendance, they are able to isolate the effect of elite schooling, which is relatively large and positive for several measures of educational attainment. Significant effects are also found for labour market outcomes (positive) and fertility (negative) in women, but not in men. Health effects of the comprehensive reform are somewhat less explored in the literature, although data shows that the distribution of health outcomes for grammar pupils strictly dominates the outcome distribution for comprehensive and secondary modern pupils(Jones, Rice, and Rosa Dias, 2012). Jones, Rice et al. (2011) implement a combination of coarsened exact and propensity score matching, before evaluating the impact of educational attainment and quality of schooling on health behaviours and outcomes. Attainment is positively associated to healthy behaviours and some health outcomes, while quality dimensions appear to be less so, controlling for cognitive skills at age 7. Finally, in a more recent study using the NCDS, Basu et al. (2018) estimate marginal treatment effects of selective schooling on health and smoking, compared to comprehensive, along the cognitive ability distribution. Using % of comprehensive schools in the individual's LEA in 1969 as a continuous instrument, they find that individuals with lower non-cognitive skills in childhood are more likely to be negatively affected by attendance to comprehensive, compared to the selective system.

In our analysis we look at two parallel questions. On the one hand, we explore long-term effects of attending grammar, compared to comprehensive, given the current political importance of assessing the potential impact of reintroducing selective schools. On the other, we investigate the effect of attending secondary modern, compared to comprehensive, since mixed-ability schools in the same areas as new selective schools would likely experience similar effects to those experienced by secondary modern schools in the past, at least in terms of pupil and resource allocation. Separating treatment effects allows us to make treatment and control groups more comparable, but we additionally implement matching as a way to pre-process the data, to achieve higher confidence that we are able to estimate an unbiased treatment effect. This is followed by parametric regressions for a rich set of outcomes. Our health outcomes include adult self-assessed physical and mental health, measured BMI and several biomarkers for risk of cardiovascular disease and stress. Well-being outcomes consist of life aspirations at the end of secondary school, adult life satisfaction, self-efficacy and job satisfaction, crime and drug use. The range of outcomes we consider is original in the literature looking at non-monetary returns of selective versus non-selective education. This broad scope allows us to build a well-rounded picture of individuals' well-being at different points of their lives after secondary school.

3 Data

The NCDS follows the lives of a cohort of individuals born in England, Scotland and Wales in a single week in March 1958. The study started out with a sample of over

[3]The 'elite' schools in the study, namely senior secondary schools, are broadly comparable to grammar schools in England, while 'non-elite' ones, known as junior secondary in Scotland, correspond to the English secondary modern.

17,000 individuals, and retained about 9,000 at the most recent wave in 2013 (Brown et al., 2016). Following the birth survey, 9 further sweeps have been undertaken to date, at ages 7, 11, 16, 23, 33, 42, 46, 50 and 55, plus the collection of biomedical samples and data at age 45. Over the years the study has collected information on cohort members' (CM) health, socio-economic circumstances, family background, education, cognitive development, beliefs, employment and familial and social relationships[4]. The key variables for the present study are described below.

3.1 Background characteristics

Detailed information on family and background available in the dataset allow us to control for a broad set of pre-schooling characteristics, responsible for the underlying differences cited as the main sources of selection bias in the estimation of the effect of schooling (Manning and Pischke, 2006). These include mother's interest in their child's schooling (expressed on a 0-4 scale), father's employment status and social class and maternal grandfather's social class, family composition, financial hardship and council accommodation tenure during childhood. Rich information is available on infant and child health, which is likely to affect both schooling and long-term health outcomes. Childhood health conditions have been previously grouped under an indicator of child morbidity (Jones, Rice, and Rosa Dias, 2011; Power and Elliott, 2006). This was constructed based on 12 morbidity categories, as classified by Power and Elliott (2006). Maternal smoking during pregnancy, presence of chronic conditions in the family (including heart, diabetes and chronic illnesses), and number of hospital admissions up to age 7 are also included to reflect health endowment. Finally, weight and height measured by a nurse at ages 7 and 11 were used to construct body mass index (BMI) measures in childhood.

3.2 Schooling

The 1958 cohort started secondary school in 1969, at a time in which the transition to the comprehensive system was still under way, meaning they experienced one of two different secondary school systems, selective and comprehensive. This setting then offers an opportunity to explore the effects of a variation in educational quality on individuals born in the same week. Information on the type of secondary school attended at age 16 is retrieved from NCDS wave 3. Schools are classed as grammar (attended by 10% of the NCDS cohort); secondary modern (20.6%); comprehensive (46.6%); non-LEA (20%), including academies, free schools, independent schools; technical (0.5%), and others (2.2%) (including all age, educationally subnormal (ESN), and other special needs). For the purpose of this paper, only the first three types are considered, leaving a sample size of 10,159 individuals going to state schools: 1,314 grammar, 2,710 secondary modern and 6,135 comprehensive pupils. The data on LEA of the school was obtained under special licence. We also manually retrieved from a 1971 edition of the Comprehensive School Committee Journal information on percentages of grammar, comprehensive and secondary modern schools in 1971 for each LEA, as well as the LEA percentage of comprehensive pupils aged 13 in 1971 (corresponding to the NCDS cohort). The latter is a more accurate measure of supply of school places than the former, and it is therefore used

[4] A detailed breakdown of the data collected for each sweep can be found in the cohort profile by Power and Elliott (2006), and in the Data Dictionary provided online by the Centre for Longitudinal Studies (www.cls.ioe.ac.uk).

to build the IV strategy for the robustness checks. Most of these percentages were supplied by LEAs at the time, while some were calculated by the CSC on the basis of school population data from the Education Committee's Yearbook of the previous academic year (Comprehensive School Committee, 1971).

Data collected at age 7 and 11 also includes information on primary school. The following is used in the empirical strategy: whether the child goes to an independent primary school; child's happiness at school, as reported by parents; whether the child will go to school or study after minimum school-leaving age, a good indicator of propensity towards education.

3.3 Ability

Cognitive skills were assessed through a range of tests, spanning numeracy, reading, verbal and non-verbal skills, during primary, just before secondary and again just after secondary school, at ages 7, 11 and 16 respectively. Of particular interest is the fact that tests administered at age 11 closely resemble the components of the 11 plus: mathematics, reading, verbal and non-verbal ability. As a further confirmation of this, Jones, Rice and Rosa Dias (2011) find that the main predictor of the propensity for attending grammar, if exposed to selection, in NCDS data, is relative cognitive ability at age 11. Following existing literature, we grouped test scores under a single indicator of cognitive ability for each wave, by implementing principal component analysis (Cawley et al., 1996, Galindo-Rueda and Vignoles, 2005, Jones, Rice, and Rosa Dias, 2011). PCA captures the variation in the data, while avoiding multicollinearity issues that would arise if all the test scores were included as regressors in the model. The factor loadings associated to the three components chosen are very similar, 0.58 each for arithmetic and general ability, and 0.56 for reading (more details in the appendix). An index constructed on the basis of these factor loadings is therefore going to mirror the 11-plus, where equal weights are given to its different components[5]. Similarly, age 7 cognitive test scores are grouped under an index by PCA. For simplicity of interpretation, we then converted the PCA index to a variable bound between 0 and 1. Finally, a rank variable is constructed from the age 11 cognitive ability index, indicating relative cognitive ability of the individual, compared to the rest of pupils in their own education system. Thus, a relative ability index ranging between 0 (lowest ability score) and 1 (highest ability score) is calculated separately for children attending the selective system (grammar and secondary modern schools) and the mixed-ability system (comprehensive schools).

Non-cognitive skills are proxied by a measure of social maladjustment, the Bristol Social Adjustment Guide (BSAG), administered at age 7 and 11. Teachers were asked to answer questions on twelve child behaviour dimensions. Lower scores indicate more developed socio-emotional, or non-cognitive, skills, while higher scores indicate more social adjustment issues and less developed non-cognitive skills. The score was also converted to a variable bound between 0 and 1 that is increasing in non-cognitive skills, and its distribution is highly skewed towards the right, due to the way the questionnaire was designed.

[5]Following Galindo-Rueda and Vignoles (2005), PCA was performed over different combinations of test scores at age 11: by aggregating all five tests available, excluding copying designs, and finally aggregating together verbal and non-verbal ability. The resulting predicted factor scores were found to be highly correlated, and therefore, in the interest of parsimony, the latter combination was used for the final age 11 ability index. Standardising the test scores to mean 0 and s.d. 1 did not change the predicted factor scores.

3.4 Outcomes

3.4.1 Health

Self-rated health (SAH) is measured on a standard 5-point scale: Excellent (1); Good (2); Fair (3); Poor (4); Very poor (5). A 9-item malaise module offers a measure of ill-health and discomfort at different ages, both physical and mental. A binary variable equal to 1 if SAH is rated excellent or good, and 0 otherwise, and a binary variable equal to 1 if malaise score is low (0,1,2), and 0 otherwise, are used as self-assessed health outcomes at age 50 for ease of interpretation. For comparability, mental health is also measured by a summary score ranging from 0 to 30 based on ten different areas: anxiety, appetite, concentration/forgetfulness, depression, depressive ideas, fatigue, irritability, panic, phobias and sleep, all measured at age 45. A body mass index (BMI) measure was constructed as (weight in kg $\div$ (height in m)^2), using measured weight and height at age 45. A healthy adult BMI ranges from 18.5 to 25kg/m^2. Individuals with smaller values would be classed as underweight, while individuals with $25 < BMI \geq 30$ would be overweight or obese if $BMI > 30$. Higher values are correlated with higher risk of cardiovascular disease, stroke and type 2 diabetes (World Health Organisation, 2017).

Blood samples taken at age 45 were used to measure lipids, clotting factors and inflammatory markers. The biometric measures we select as outcomes include C-Reactive protein (CRP) levels (g/L), fibrinogen levels (g/L) and triglyceride levels, as well as cholesterol ratio, constructed as (total cholesterol$\div$HDL cholesterol). All of these markers are positively linked to risk of cardiovascular disease (Benzeval et al., 2014). CRP and fibrinogen are also associated with higher risk of chronic stress, bringing together the physical and mental domains. The use of biomedical outcomes represents an original element of our study in the literature on the effects of school quality, as it allow us to characterise health status on a finer scale. In addition to present health problems, we are thus able to assess the probability of health problems arising in the future.

3.4.2 Well-being outcomes

Individual well-being is a complex concept, including not only health and wealth, but much more. In order to assess short-term impact of secondary schooling, we look at aspirations related to school and work at age 16, potential determinants of future achievements, measured just after secondary school. School aspirations is a variable constructed via PCA that groups cohort members' answers to five related questions, measuring the individual's attitude towards school and studying. Work aspirations indicates whether the individual aspires to personal and intellectual growth through a job. The outcome measure is a dummy variable equal to 1 if the individual rates 'Using head', 'Involves variety' and 'Good prospects' among their top three priorities in terms of job attributes. Life satisfaction, self-efficacy and positive feelings about one's job, based on the age 33 survey, are all constructed via PCA, grouping answers to several questions. Finally, contact with police and drug use are retrieved at age 45. The crime dummy indicates whether the individual had any significant contact with police (i.e. whether ever moved by police, received a warning, got arrested, cautioned, or found guilty). The dummy for drugs takes value 1 if the individual has ever tried any of 12 types of drugs, or any other illegal drug[6]. Given the smaller health risks connected to the infrequent use of cannabis,

[6] A more meaningful indicator would be a variable measuring individual's frequency of use or dependence. Such information was available, but it was not used as it only concerned a very small percentage

this is excluded from the analysis.

4 Methods

For our model we draw from the framework of individual investment in their human capital, represented here by health and other well-being dimensions. We assume three time periods t =*0,1,2*, corresponding roughly to infancy, childhood (just prior to secondary school entry) and adulthood. Individuals start out with background characteristics B_{i0}, comprising family and individual characteristics, health endowment and socio-economic status. Background and a genetic endowment of cognitive and non-cognitive abilities A_{i0} determine ability prior to entrance into secondary school:

$$A_{i1} = A(B_{i0}, A_{i0}) \tag{1}$$

Secondary school assignment S_{i1} (i.e. school type) is the key treatment of interest, and we assume it is also a function of pupil's background and childhood abilities (this is particularly true in selective areas), as well as characteristics of the individual's LEA, such as supply of places by type of school.

$$S_{i1} = S(B_{i0}, A_{i1}(.), SU_{i1}) \tag{2}$$

The production function for health and well-being outcomes, Y_{i2}, depends on background, pre-secondary school ability, type of school (our treatment of interest), and local area characteristics.

$$Y_{i2} = Y(B_{i0}, A_{i1}(.), S_{i1}(.), LA_{i1}) \tag{3}$$

Note that in our model we exclude any post-treatment variables, as these might bias the treatment effect in the empirical estimation. For the same reason health behaviours adopted in adulthood are not included in the empirical specification either. In the model, background B_{i0} and ability A_{i1} enter both the school-assignment function, (2), and the outcome equation of interest, (3). If the relevant variables in the empirical specification do not capture all the relevant dimensions of background and ability, then, given they enter a third covariate, treatment S_i, the estimated coefficient on S_i will be biased. This issue represents the main challenge for identification of treatment effect in this context. The main aim of the present work consists of implementing an appropriate strategy in order to isolate the effect of S_{i1} on the outcomes of interest. As pointed out above, selection bias has been a concern in the literature using the NCDS, given that children attending schools in comprehensive and selective areas might differ in both observable and unobservable characteristics before they start secondary schooling (Manning and Pischke, 2006, Bonhomme and Sauder, 2011).

Isolating treatment effects to establish more than simple correlations requires comparing treated individuals with credible counterfactuals (Rubin, 1974, Heckman et al., 1997). In this spirit, we split the sample into two, and aim at estimating two separate treatment effects. On the one hand, we estimate the effect of going to grammar, compared to comprehensive, by comparing outcomes for grammar pupils to those of comprehensive pupils who would have gone to grammar, had they gone through selection.

$$ATT^G = E[Y_i^1 - Y_i^0 | G_i = 1] \tag{4}$$

of the total sample.

Similarly, we estimate the effect of going to secondary modern, compared to comprehensive, by contrasting secondary modern pupils and comprehensive pupils who would have attended secondary modern, had they experienced the selective system[7].

$$ATT^{SM} = E[Y_i^1 - Y_i^0 | SM_i = 1] \tag{5}$$

The way we ensure we compare like with like is via matching methods, aimed at increasing balance in observable baseline characteristics between the treatment and control groups (Angrist, 1998). This first step is followed by parametric regressions based on the model expressed by Equation 3, and estimated using the weights obtained in the matching procedure. Our empirical approach therefore develops in two steps: matching, as a form of pre-processing the data, is combined with parametric regressions. The latter rely on a set of assumptions, such as the functional form used and the specification of variables included in the model. While these are justified on the grounds of economic theory and previous established literature, reliance on these assumptions can be seen as a weakness of the empirical analysis. This is particularly the case where there is a lack of common support across treated and control units[8]. Then, matching on raw data and using resulting weights in subsequent parametric approaches can help reduce model dependence on crucial, although not entirely verifiable, assumptions (Ho et al., 2007). The advantage of this approach is that it yields 'doubly robust' estimates: treatment effects will be consistently estimated if the matching achieves balance, even though subsequent parametric models are not well specified; or if matching is incorrect, while parametric models are well specified (Ho et al., 2007).

4.1 Preprocessing data: matching

Matching is implemented separately for the two samples. The first sample includes grammar and comprehensive pupils (GC sample hereafter), with grammar school attendance as treatment. The second comprises secondary modern and comprehensive pupils (SMC sample hereafter), with secondary modern attendance as treatment. The idea is to make comprehensive pupils a credible counterfactual group in each of the two instances where we estimate treatment effect. For example, we expect the comprehensive matches to grammar pupils to display higher average cognitive ability scores than secondary modern pupils and their respective comprehensive matches. Upon surveying a range of matching procedures, entropy matching was found to achieve the best balance among the covariates of interest, while retaining all important information from the original sample[9]. Developed by Hainmueller (2012), entropy matching assigns weights to the observations in the control group according to pre-specified conditions, in order to achieve balance on the moments and co-moments of specific covariates. The Stata package ebalance allows for a straightforward implementation of the method, which is another benefit of the present procedure.

[7]Another way to look at it is as the effect of going to high-ability school versus all-ability school, and the effect of going to low-ability school compared to all-ability.

[8]Common support holds when for each value of a given covariate X, $0 < P(S = 1|X) < 1$

[9]Alternatives surveyed included propensity score matching, and a combination of coarsened exact matching followed by propensity score matching (Iacus et al., 2011, Leuven and Sianesi, 2012). Although the quality of the matches was lower, and the sample size reduced due to observations outside commmon support being dropped, results in the outcome regressions are not significantly dissimilar following the three alternative matching strategies.

The candidate covariates for matching are selected on the basis of whether they are expected to be related to both treatment and outcomes (Caliendo and Kopeinig, 2008)[10]. Childhood cognitive skills, socio-emotional ability, and parents' interest in child education and socio-economic status are all deemed to be key determinants of treatment assignment, as well to affect future outcomes. In order to ensure that the variables are not influenced by treatment, which would bias effect estimates, only pre-secondary schooling variables are used. Still, some such variables could be affected by the anticipation of treatment; this might be the case for cognitive ability scores at age 11 if there are coaching effects[11], so only age 7 cognitive ability scores are selected for matching (Jones, Rice, and Rosa Dias, 2011). Age 11 relative position by cognitive ability, on the other hand, is included, as well as age 11 BSAG score, our non-cognitive skills indicator[12]. As background variables we include mother's interest in child education, measured on a scale from 0 to 4, and a dummy for high or middle-high father's SES, both measured when the child is aged 11. We decided to balance mean, variance and skewness of the included covariates, as well as their pairwise interactions, as we thus achieved very close balance, without compromising the feasibility of the minimization procedure[13].

4.2 Parametric modelling

In contexts where treatment assignment can reliably be predicted as a function of observables, matching estimators yield a consistent estimate of treatment effect. This key assumption, conditional independence, is expressed as

$$Y_i^j \perp S_i | \mathbf{X}_i, \tag{6}$$

with $j = 0, 1$. The above states that conditional on the vector of observable characteristics $\mathbf{X}_i$, selection into treatment S_i does not depend on potential outcomes. In the present case however, we are not interested in treatment effect directly estimated via matching. Rather, we use the weights obtained via entropy matching for the control units in a parametric regression where we control for all available covariates that might affect our outcomes of interest, in addition to the five key covariates used in the matching procedure. In the choice of the empirical strategy for estimation of Equation (3), we start with the prior that treatment assignment S_i could be endogenous, a classic problem in the literature on returns to education[14]. Endogeneity can arise for different reasons, examples being reverse causality and/or omitted variables, common when there are unobserved confounders in the relationship between education and outcomes. Since OLS estimation

[10]The methodological literature highlights that the choice of which covariates to match on yields a trade-off between bias and efficiency (Rubin and Thomas, 1996, Imbens, 2004). Matching on a variable that is related to treatment but not outcome will increase variance of the effect estimate; conversely, matching on a covariate related to outcome but only weakly to treatment will bias the estimate.

[11]Coaching effect is the term used in the literature for the idea that students in selective LEAs might score higher in ability tests at age 11 because they have been coached to pass tests of that specific type, in view of the imminent 11-plus exam (Jones, Rice, and Rosa Dias, 2011).

[12]Since the rank variable is constructed separately for selective (grammar and secondary modern school) and non-selective (comprehensive) pupils (see Section 3.3), the bias of coaching effects does not apply.

[13]More details on entropy matching and the ebalance package in Stata available in Hainmueller (2012).

[14]For any given outcome of interest

$$Y_i^j = \mathbf{X}_i \beta + \epsilon_i, \tag{7}$$

any element of $\mathbf{X}_i = (S_i, X_{1i}, ..., X_{Ni})$ correlated to the error term ϵ_i is said to be endogenous.

in the presence of endogeneity yields biased and inconsistent estimates, it is important to conduct some checks in order to find the estimation method most likely to avoid bias.

We conduct Durbin-Wu-Hausman tests of endogeneity of school type for all outcomes of interest, before and after matching, including all available controls[15]. For all our outcomes, the test fails to reject the null of exogeneity (full test results available on request). Under treatment exogeneity, OLS has superior finite sample properties to IV estimators, as well as smaller variance, and therefore we proceed to estimating Equation (3) by OLS as our main empirical strategy (MacKinnon and Davidson 2003). Our hypothesis is that the rich set of control variables available, including measures of different abilities, and a broad range of socio-economic and local area characteristics, allow us to control for all the main confounders in the relationship between our treatment and outcomes.

4.3 Robustness checks

As already stated, the Durbin-Wu-Hausman test does not reject exogeneity and we use OLS as our primary estimator. Nevertheless, as a robustness check, we implement IV strategies after preprocessing the sample via matching. As mentioned, the literature has used share of comprehensive schools in the LEA at time of schooling (Basu et al., 2018), and LA political control (Galindo-Rueda and Vignoles 2005). Here we use percentage of 13-year-old pupils going to comprehensive schools in each LEA in 1971, retrieved by the Comprehensive School Committee 1971 Journal, as main IV to instrument type of school attended by NCDS cohort members[16]. The choice of instrument, Z_i, is justified according to the following criteria. First, since % of comprehensive pupils in a given LEA is a measure of supply of comprehensive places, both grammar and secondary modern attendance are expected to be significantly and negatively correlated with the instrument, therefore $corr(Z_i, S_i) \neq 0$. Secondly, the validity of the exclusion restriction assumption requires the instrument to be exogenous with respect to the outcome of interest, $cov(Z_i, \epsilon_i) = 0$, where ϵ_i is the error term in the outcome Equation (3). The large size of the LEAs (average LEA population in 1971 was 413,649 (Registrar General for England and Wales, 1971)) makes it unlikely that LEA-level characteristics could determine individual's behaviours and outcomes as much as, say, neighbourhood- or school-related characteristics could, as peer effects and environmental factors are weaker in larger areas. As a further check of this, several LEA characteristics, such as county proportion of unemployed, council tenants, house owners and professional categories for household heads, are regressed on the instrument and other individual characteristics. Unfortunately, correlation between the instrument and these other LEA-level characteristics is in some cases significant, which weakens the validity of the instrumental strategy. In order to minimise this problem, these LEA-level characteristics are controlled for in the parametric specification of all outcome regressions.

[15]The DWH test allows testing for endogeneity in just-identified models. For each outcome, the null hypothesis is that the treatment variable is exogenous. The residuals from the first stage of the 2SLS procedure are included as a regressor in the outcome regression with the original (not the predicted) treatment variable. If first-stage residuals are not significantly associated with the outcome, then this is taken as evidence for treatment exogeneity. Since our model is just-identified we cannot implement the Sargan J test, which requires over-identification (i.e. more than one instrument for one treatment).

[16]Percentage of grammar schools in the LEA as a share of total schools and share of secondary modern schools were available, but not as precise as share of pupils when it comes to proxying school supply, which is the rationale for using this type of variable as an IV. Percentage of grammar and secondary modern pupils in each LEA was not available from the sources mentioned.

The models for the outcomes of interest are estimated by Two Stage Least Squares (2SLS). Two Stage Residual Inclusion (2SRI) methods, allowing for non-linear models in either the first or second stage or both, are also explored as an alternative, but results are left in the appendix. The first stage of 2SLS, the empirical counterpart of Equation (2), consists of the school assignment function, using percentage of comprehensive pupils in the individual's LEA as an instrument. In 2SLS, the second stage uses the school type predicted in the first as a regressor for the outcome equation. This second stage is the empirical counterpart to Equation (3). In theory, 2SRI has the advantage that it allows non-linear specification of either or both of the first and second stage regression (Terza et al., 2008). In the present case, this means that the first stage can be expressed as a probit model, which is a better fit for a binary outcome such as school assignment. Residuals from this first stage[17] are then plugged in as a regressor in the second stage, which is expressed as a linear or probit model, depending on the specific outcome[18]. Importantly, the original endogenous regressor, and not the one estimated in the first stage, is included in the second stage of 2SRI. Terza et al., 2008 show that this method consistently estimates treatment effects, based on the idea that residuals from the estimation of the first stage capture the unobserved confounders causing the endogeneity problem. Note that in all of our estimations, residuals from the first stage were never significant predictors of the outcome, which is essentially confirming that we cannot reject exogeneity of school type, as already supported by the Durbin-Wu-Hausman test.

Given the NCDS cohort entered secondary school in 1969, only four years after the Labour-backed Circular 10/65, we conduct a further check distinguishing between comprehensive schools that were formerly grammar or secondary modern, versus comprehensive that are either purpose-built or amalgamated. So our new specification for this robustness check within the GC sample will have grammar as the base category and two treatment variables: one being an indicator for attendance to a comprehensive that is a former grammar, and one indicating attendance to any other type of comprehensive. A symmetric approach is then implemented to the SMC sample too.

We further include interaction of the treatment and ability variables in order to explore heterogeneity of treatment effect along the cognitive and non-cognitive ability distributions. We explore interactions of school type with ability quartiles. The estimated coefficient will then reflect the effect of grammar attendance, say, compared to comprehensive, for pupils in the highest 25% of the ability distribution and so on.

A further robustness check of the approach used follows the placebo test procedure implemented by Manning and Pischke (2006), in order to increase confidence in our identification strategy. Essentially, the procedure consists of estimating the effect of type of secondary school for two models, one with post-secondary school maths test scores as the outcome, and one with pre-secondary school maths test scores. In theory, we would not expect secondary school type to be a significant predictor of pre-secondary school scores, unless the model is misspecified or the estimation strategy unable to prevent bias. In some recent work, Basu et al. (2018) conduct a similar check, taking child morbidity as main outcome of the placebo procedure. In our falsification tests, conducted separately for the GC and SMC samples, we first take the same outcomes used by Manning and Pischke (2006), which are maths score administered at ages 16 and 11, and then BMI

[17]The calculation of generalised residuals from the probit model is left in the appendix.

[18]Note that 2SLS estimates the second stage as a linear probability model for all outcomes, including binary ones. While this is often accepted in the interest of parsimony (Angrist and Pischke, 2009), 2SRI allows us to relax linearity in a convenient way.

at ages 11 and 16, as a measure of general health[19]. If the matched sample passes the falsification test, this will strengthen the hypothesis that our two-step procedures are able to identify the effect of interest, and that this is not due to pre-existing differences in the compared samples.

5 Results

5.1 Characteristics by type of school

Table 1 describes the main individual characteristics of interest, by type of school. All are measured when the child is aged 11, unless otherwise specified. Future grammar, comprehensive and secondary modern pupils differ most notably in the three measures of ability. At 11, grammar pupils present highest cognitive ability, the main determinant of entry test success, followed by comprehensive and then by secondary modern pupils. Age 7 and 16 cognitive scores also follow a similar pattern. Grammar pupils also display higher non-cognitive abilities. Grammar pupils present slightly higher proportions of female and first-born children, and are less likely to have two or more siblings. The figures for the family background variables suggest that on average grammar pupils are more advantaged, both in terms of parental interest in their education and social and financial situation. Grammar pupils are much more likely to state at 11 that they plan to study after compulsory schooling. They have slightly higher health endowment than the other two categories of pupils, with lower probabilities of mothers smoking during pregnancy and presence of a chronic illness in the family. Average local area characteristics, as registered in the 1971 census, are very similar across the three groups, somewhat reassuringly for the identification of unbiased treatment effects. The only notable exception, as expected, is the externally retrieved instrumental variable, percentage of comprehensive pupils as a share of total pupils in the individual's LEA, which is highest for comprehensive pupils, compared to the rest of the sample.

As shown in Table 2, on average grammar pupils display better health for all outcomes considered, while these are worst for secondary modern students out of the three groups. Lower values of self-assessed health denote better health, and so do malaise and mental health score values. All biometric measures are increasing in bad health and risk of cardiovascular disease and diabetes, as well as stress. On average, grammar pupils score lowest in all the biomarkers considered, while secondary modern present the highest risk in all these measures. Well-being outcomes at different ages present similar trends: grammar pupils generally do better at all ages than the other two groups, while secondary modern do worst, except for life satisfaction, where comprehensive pupils score highest and grammar pupils score lowest. At 45, grammar pupils are less likely to have had significant contact with police, but slightly more likely to have tried an illegal drug.

5.2 Main results

Tables 3 and 4 show the first three moments of the five key covariates of interest, before and after entropy matching, for the GC and SMC samples respectively. The lower panel in both tables shows re-weighted mean and variance for the comprehensive sample, based

[19]BMI is preferred here to a more general measure of health because of its simplicity, and the difficulty of finding comparable measures of health across the first waves of the NCDS.

on the entropy balancing weights. In both samples almost perfect balance is achieved on mean, variance and skewness of key covariates for the treated and control groups. The pairwise interactions between covariates are not shown, but almost perfect balance is also achieved on their moments, increasing confidence that the joint distribution of these variables will be more similar in the two groups after matching.

Tables 5 to 8 report results for the main outcome regressions of interest, all estimated by OLS for the matched GC and SMC samples separately. Although we only show the coefficients on the treatment and a few more key variables, all models are estimated controlling for all covariates described in Table 1 (full results available upon request from the authors). Table 5 shows that grammar attendance does not affect long-term health outcomes, compared to comprehensive, with the exception of BMI, for which the effect is negative and only significant at 10%. The ability variables, on the other hand, are all significant for at least some outcomes. Cognitive ability at age 7 is negatively and significantly related to age 45 BMI levels, as well as to biomarkers for risk of CVD. More specifically, a 0.10 increase in the cognitive ability index is associated with a reduction of 6% of a standard deviation (s.d.) in BMI and a reduction of 5% and 6% of a s.d. in cholesterol ratio and tryglicerides respectively. Higher socio-emotional skills are only associated with lower incidence of mental health problems in the grammar and comprehensive sample. Relative position of the child by cognitive ability at age 11 is also predictor of CRP and fibrinogen levels. An increase of 10 percentage points in relative position (sufficient to be shifted to the next upper decile of the cognitive ability distribution), is associated with s.d. decreases of 38% and 1% respectively in the risk biomarkers. Ceteris paribus, women have poorer mental health, with lower probability of scoring low on the malaise scale, and higher incidence of mental health problems. However, their BMI is on average 1 point lower and they score lower in all biomarkers for risk, except for CRP .

Table 6 shows well-being outcome results for the matched grammar and comprehensive pupils. Grammar is significantly related to increases in positive aspirations about school and studying at age 16 (equivalent to 22% s.d.), but with lower life satisfaction at age 33 (magnitude 12% s.d.), compared to comprehensive attendance. Again cognitive and non-cognitive abilities present striking results. A 0.10 increase in non-cognitive skills is associated with higher positive feelings about school at 16 (8% s.d.) and higher life satisfaction and self-efficacy at 33 (9% s.d. and 12% s.d. respectively), as well as drug use. Relative cognitive ability is significantly linked to higher positivity about school and work at age 16, as well as higher self-efficacy and job positivity at 33. Magnitudes of these effects for a 10 percentage point increase in relative cognitive ability are 16%, 7%, 4% and 13% s.d. of the respective outcome. Interestingly, relative cognitive ability is also significant at 10% as a predictor of higher drug use.

Table 7 shows results for health outcomes, estimated for the matched secondary modern and comprehensive sample. Secondary modern attendance is not a significant predictor for any of the health outcomes considered, compared to the alternative of going to comprehensive. Again, non-cognitive skills and relative cognitive ability at age 11 are significantly related to later health outcomes. A 0.10 increase in the non-cognitive skills score is linked to increases in the probabilities of scoring high self-assessed health and low malaise at age 50, as well as a reduction of 7% of a s.d. in mental health problems at 45. Relative position by cognitive ability is also related to increases in both these probabilities, as well as lower fibrinogen levels (magnitudes of effects of 3%, 10% and 6% of a s.d. for each outcome respectively, for a 0.10 increase in relative position). An interesting difference in this second SMC sample, compared to the GC one, is that now mother's interest and father's SES seem to account for some of the variation in health. Notably this is the case with BMI, which is negatively related to mother's interest in child education and high paternal SES in the SMC sample, but not the GC sample.

Finally, Table 8 shows results for well-being outcomes, estimated using the matched secondary modern and comprehensive sample. Secondary modern attendance increases positive feelings about school at 16 (12% s.d.) and self-efficacy at 33 (8% s.d.), compared to comprehensive. Pre-secondary schooling non-cognitive skills are once more a significant predictor of well-being, and they are positively and significantly related to all of our positive outcomes, while negatively related to contact with police and drug use. Effects range between 3% of a s.d. (for work aspirations at 16) and 6% of a s.d. (for life satisfaction at 33) for a 0.10 increase in non-cognitive skills. Relative cognitive ability is again positively and significantly related to well-being: a 0.10 increase in relative position is related to increases in well-being of between 2% and 14% of a s.d., depending on the outcome.

5.3 Robustness checks

Tables 9 to 13 display results for 2SLS estimation of the IV models used as a robustness check for our main results. Table 9 presents results for the first stage of 2SLS estimation, the key variable of interest being the percentage of comprehensive pupils in the individual's LEA, which is the IV of choice. The instrument is a significant and negative predictor of grammar and secondary modern attendance for each sample respectively, and the overall F-test is always greater than 10, which by rule of thumb increases the confidence that the instrument of choice is not weak, thus strengthening the credibility of this strategy to identify school assignment[20] (Stock et al., 2002).

Our four 2SLS results tables show that the few significant effects of grammar and secondary modern found with OLS estimation are not corroborated by our IV models. On the other hand, 2SLS coefficients for the ability variables in all four of our IV results tables are largely similar to OLS ones in magnitude and significance. 2SRI results, here not shown, are similar to OLS results, although the magnitude of coefficients varies in several cases. Additionally, as noted above, generalised residuals saved from the first stage are never significant, indicating either that the term is unable to capture unobserved confounders in the structural equation, or that endogeneity in this instance is not a problem.

Considering the effect of comprehensive schools that were formerly grammar or secondary modern separately from other types of comprehensive schools did not offer any further interesting insights. Interacting treatment with high and low levels of cognitive and non-cognitive skills did not produce significantly different results either (see appendix tables).

5.4 Falsification test

As a further check, we conduct placebo procedures in the same spirit of Manning and Pischke's falsification test, in order to support feasibility of the empirical strategies presented in Section 4 (results are left in the appendix). There are some key differences to

[20]Note that in just-identified models (i.e. where there is one instrument for each endogenous variable), weak instrument bias is much smaller than in over-identified ones, especially if the first stage is highly significant.

note in our procedure, compared to Manning and Pischke's original approach: first, we implement the regressions separately for the GC and SMC samples, instead of considering the whole sample; second, since we are interested in health outcomes, along the lines of Basu et al. (2018) , we add BMI at ages 11 and 16 as an alternative outcome to maths scores; third, we include our own set of control variables, as in our main outcome regressions. The idea is that if the matched samples with parametric regressions 'pass' the falsification test, meaning that the coefficient on comprehensive is not significant for age 11 outcomes, then this goes to support our identification procedure for the effect of school type.

Following the original paper, maths test scores are converted on a scale from 0 to 100, so that they are more easily interpreted. The score 16 and 11 results confirm what was found by the authors. Comprehensive attendance, used as treatment for both groups for comparability with the original test, is a significant and negative predictor of both age 16 and age 11 maths scores for the GC sample. For the SMC sample, comprehensive attendance is a positive and significant predictor of maths scores at age 11, while insignificant for age 16. This puzzling result might be explained by coaching effects: future secondary modern pupils, who were more likely to live in areas with grammar schools, were also more likely to be exposed to coaching for the 11-plus, which would increase their maths test scores at age 11 compared to their counterparts living in comprehensive areas, without necessarily indicating higher cognitive abilities. This difference would then be eliminated at age 16.

Since our primary outcomes of interest are health and well-being, we carry out similar procedures with BMI at age 16 and 11 as dependent variables. Note that we add non-cognitive skills as a further ability variable, to follow more closely our main identification strategy. Comprehensive attendance is never significant for age 16 nor age 11 BMI, in either sample. Both samples show some significant associations between BMI and cognitive and non-cognitive ability. For both samples, lower age 11 non-cognitive ability scores are linked to increases in BMI at age 16. These results strengthen credibility of matching as a way to pre-process the data before implementing parametric regressions for health outcomes, while they suggest some caution in using a similar strategy for education outcomes. Moreover, what we find in terms of childhood and adolescence BMI confirms the results of the main outcome regressions: school type is not a key determinant of long-term health, while childhood cognitive and non-cognitive abilities prior to secondary schooling are.

6 Discussion

Our paper adds to the literature on the effects of selection by ability in secondary schooling, specifically in relation to England and health and well-being outcomes. A key contribution of our paper is the combination of covariate balancing via matching methods with parametric regression, which yields doubly robust estimates, thus helping to address the problem of endogeneity of schooling. The previous criticism advanced by Manning and Pischke (2006) was aimed at value-added methodologies and IV regressions, which were used to explore the effects of selective schooling on educational achievement and later shown to be unable to eliminate selection bias. Our placebo procedures, in the same spirit as theirs, confirm that our methodologies are able to deal with selection bias when estimating models of health outcomes, while we should be cautious about drawing implications for education outcomes.

Our findings corroborate previous literature on the effects of selective schooling. With different methods from those used by Bonhomme and Sauder (2011), we also find that when correcting for pre-treatment differences in pupil characteristics, the average effect of type of school is not a significant predictor of long-term outcomes. Basu et al. (2018) reach a very similar conclusion for health and smoking, and we extend their result by adding evidence on biomarkers for risk of CVD and well-being outcomes. However, when exploring heterogeneity by looking at person-centered treatment effects of selective schooling, the same authors find a significant and persistent health effect for individuals with a set of characteristics - specifically men with lower childhood non-cognitive skills. Our consideration of average treatment effect only is then a limitation of our analysis, as it could hide heterogeneous dynamics. Heterogeneity analysis could inform us on whether there are characteristics to which a larger or more significant treatment effect is associated, which could have meaningful implications for educational policies. Similarly to Jones, Rice and Rosa Dias (2011), we find that even accounting for several underlying differences, non-cognitive abilities are important predictors of long-term outcomes, and extend this result to our health outcomes. We also find that cognitive abilities can be important determinants of health, even when accounting for non-cognitive ones, which agrees with literature in the field (Auld and Sidhu, 2005, Conti and Heckman, 2010, Bijwaard, Kippersluis, et al., 2015).

This suggests that a reason why we do not find a significant treatment effect is that secondary schooling might not affect the channels that are assumed to lead to better health and higher well-being. For instance, cognitive and non-cognitive skills, as well as preferences determining our decisions, might be shaped earlier on in childhood. If it is true that they affect health and well-being in the long-term, then educational policy might have larger spill-over effects on health if it channels its budget towards early childhood education interventions, rather than new selective schools. On the other hand, channels that affect health might also be formed later on, after secondary schooling. This is the case of changes produced in adulthood via University attendance, career path, work environment and so on. Research on the mediators between education and health and on the key life stages at which these mediators are affected could inform us more on the potential beneficial effects of educational policy for health and well-being (Kautz et al., 2014). Preliminary analysis using our data showed us associations between type of school attended and cognitive ability at age 50. Cognitive scores and social integration were found to be important mediating pathways in the relationship between education and health, when estimating the effect of achieving A-levels on self-assessed health and health behaviours in the NCDS cohort, although no causal claim was made (Cutler and Lleras-Muney, 2011). Bijwaard and Jones (2016) investigate the role of cognitive ability in the relationship between education and mortality as either selection variable determining treatment, or mediator, and find that at lower education levels, differences in mortality are traced back to differences in cognitive ability.

7 Conclusion

We add a timely piece of evidence to the current debate on the reintroduction of selective schools in England, by looking at long-term health and well-being effects of making it into grammar school or being left out, compared to going to a mixed-ability school. We use data from the 1958 British birth cohort, whose members attended both a selective system, separating children by ability at age 11 into different schools, and a mixed-ability, or comprehensive, system, allowing us to explore health and well-being effects at several points of these individuals' lives over time. Critics have argued that the literature that has looked at the effect of type of school for this cohort has been unable to eliminate the standard selection problem characterising the analysis of returns on education (Manning and Pischke, 2006, Bonhomme and Sauder, 2011). We address these criticisms, following previous work by Jones, Rice and Rosa Dias (2011), by implementing a two-step empirical procedure to identify treatment effects robustly. The data is first preprocessed through a combination of matching methods, followed by parametric regression analysis.

Our findings suggest that there is no long-term direct impact of high- or low-ability school attendance compared to mixed-ability school attendance on self-assessed health, risk for cardiovascular disease, risk of chronic stress and well-being measures at different ages. Childhood cognitive and non-cognitive ability measured prior to secondary schooling, on the other hand, play a significant role as predictors of later health. Their role as either direct causal predictor of health or mediators between education and health should be the subject of further research to explore the determinants of individuals' health investment decisions. Findings could have interesting implications in terms of educational policy, which could integrate these skills as part of the school curriculum or targeted skill-boosting programmes.

8 References

Albouy, Valerie and Lequien, Laurent (2009). "Does compulsory education lower mortality?" In: *Journal of Health Economics* 28.1, pp. 155–168 (cit. on p. 1).

Angrist, Joshua D (1998). "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants". In: *Econometrica* 66.2, pp. 249–288 (cit. on p. 10).

Arendt, Jacob Nielsen (2005). "Does education cause better health? A panel data analysis using school reforms for identification". In: *Economics of Education Review* 24.2, pp. 149–160 (cit. on p. 1).

Auld, M. Christopher and Sidhu, Nirmal (2005). "Schooling, cognitive ability and health". In: *Health Economics* 14.10, pp. 1019–1034 (cit. on p. 18).

Ayyagari, Padmaja, Grossman, Daniel, and Sloan, Frank (2011). "Education and health: Evidence on adults with diabetes". In: *International Journal of Health Care Finance and Economics* 11.1, pp. 35–54 (cit. on p. 1).

Basu, Anirban, Jones, Andrew M., and Rosa Dias, Pedro (2018). "Heterogeneity in the impact of type of schooling on adult health and lifestyle". In: *Journal of Health Economics* 57, pp. 1–14 (cit. on pp. 2, 12).

Benzeval, Michaela et al. (2014). "Biomarker User Guide and Glossary". In: *Understanding Society: The UK Household Longitudinal Study*, pp. 2010–2012 (cit. on p. 8).

Bijwaard, Govert E and Jones, Andrew M (2016). "Cognitive Ability and the Mortality Gradient by Education : Selection or Mediation ?" In: *IZA DP No. 9798 Cognitive* 9798.

Bijwaard, Govert E, Kippersluis, Hans Van, and Veenman, Justus (2015). "Education and health : The role of cognitive ability". In: *Journal of Health Economics* 42, pp. 29–43 (cit. on p. 18).

Bolton, Paul (2016). *Grammar School Statistics* (cit. on p. 3).

Bonhomme, Stéphane and Sauder, Ulrich (2011). "Recovering distributions in difference-in-differences models : a comparison of selective and comprehensive schooling". In: *Review of Economic Studies* 93.May, pp. 479–494 (cit. on pp. 2, 9, 19).

Brown, Matt, Dodgeon, Brian, and Mostafa, Tarek (2016). *Webinar: Introduction to the National Child Development Study* (cit. on p. 6).

Caliendo, Marco and Kopeinig, Sabine (2008). "Some practical guidance for the implementation of propensity score matching". In: *Journal of Economic Surveys* 22.1, pp. 31–72 (cit. on p. 11).

Campbell, Frances et al. (2014). "Early Childhood Investments Substantially Boost Adult Health". In: *Science* 28.3436178, pp. 1478–1485 (cit. on p. 1).

Cawley, John et al. (1996). "Measuring the Effects of Cognitive Ability". In: *NBER Working Paper Series* 5645.July (cit. on p. 7).

Comprehensive School Committee (1971). *Comprehensive Education: bulletin of the Comprehensive Schools Committee.* (Cit. on p. 7).

Conti, Gabriella and Heckman, James J. (2010). "Understanding the Early Origins of the Education– Gradient: A Framework That Can Also Be Applied to Analyze Gene–Environment Interactions". In: *Perspecti Psychol Sci.* 5.5, pp. 585–605 (cit. on p. 18).

Cutler, David M. and Lleras-Muney, Adriana (2011). "Understanding Differences in Health Behaviors by Education". In: *Journal of Health Economics* 29.1, pp. 1–28 (cit. on pp. 1, 18).

Del Bono, Emilia and Clark, Damon (2014). “The Long-Run Effects of Attending an Elite School: Evidence from the UK”. In: *Institute for Social and Economic Research Working Paper No. 2014-05* 8.1, pp. 150–176.

Fuchs, Victor R. (1982). “Time Preference and Health: An Exploratory Study”. In: *NBER Chapters* I, pp. 93–120 (cit. on p. 1).

Galindo-Rueda, Author Fernando and Vignoles, Anna (2005). “The Declining Relative Importance of Ability in Predicting Educational Attainment”. In: *The Journal of Human Resources* 40.2, pp. 335–353.

Galindo-Rueda, Fernando and Vignoles, Anna (2005). “The Heterogeneous Effect of Selection in Secondary Schools: Understanding the Changing Role of Ability.”

Grossman, Michael (1972). “On the Concept of Health Capital and the Demand for Health”. In: *Journal of Political Economy* 80.2, pp. 223–255 (cit. on p. 1).

Hainmueller, Jens (2012). “Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies”. In: *Political Analysis* 20, pp. 25–46.

Heckman, James J., Smith, Jeffrey, and Clements, Nancy (1997). “Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts”. In: *The Review of Economic Studies* 64.4, pp. 487–535 (cit. on p. 9).

Ho, Daniel E et al. (2007). “Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference”. In: *Political Analysis* 15, pp. 199–236 (cit. on pp. 2, 10).

Iacus, Stefano M, King, Gary, and Porro, Giuseppe (2011). “Causal Inference without Balance Checking : Coarsened Exact Matching”. In: *Political Analysis* (cit. on p. 10).

Imbens, Guido W (2004). “Nonparametric estimation of average treatment effects under exogeneity: a review”. In: *Review of Economics & Statistics* 86.1, pp. 4–29 (cit. on p. 11).

Jesson, David (2000). “The Comparative Evaluation of GCSE Value-Added Performance in Two-Sector Models of Endogenous Growth by Type of School and LEA”. In: *University of York Discussion Papers in Economics* 2000.52.

Jones, Andrew M, Rice, Nigel, and Rosa Dias, Pedro (2011). “Long-Term Effects of School Quality on Health and Lifestyle : Evidence from Comprehensive Schooling Reforms in England”. In: *Journal of Human Capital* 5.3, pp. 342–376 (cit. on pp. 2, 6, 7, 11).

Jones, Andrew M, Rice, Nigel, and Rosa Dias, Pedro (2012). “Quality of schooling and inequality of opportunity in health”. In: *Empirical Economics*, pp. 369–394 (cit. on pp. 2, 5).

Kautz, Tim et al. (2014). *Fostering and Measuring Skills : Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success.* Tech. rep. 110. OECD, pp. 1–87 (cit. on p. 18).

Kerckhoff, Alan (1986). *Efffects of Ability Grouping in Secondary Schools in Great Britain* (cit. on p. 2).

— (1996). *Going Comprehensive in England and Wales: a study of uneven change.* London: Woburn Press (cit. on p. 2).

Leuven, Edwin and Sianesi, Barbara (2012). *PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing* (cit. on p. 10).

Long, By Robert, Foster, David, and Roberts, Nerys (2017). “Grammar schools in England: Briefing Paper 7070”. In: 7070, pp. 1–28 (cit. on p. 3).

Manning, Alan and Pischke, Jörn-Steffen (2006). "Comprehensive Versus Selective Schooling in England in Wales: What We Know?" (Cit. on pp. 2, 6, 9, 19).
Oreopoulos, Philip (1997). "Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter". In: *American Economic Review* 96.1, pp. 152–175 (cit. on p. 1).
Power, Chris and Elliott, Jane (2006). "Cohort profile: 1958 British birth cohort (National Child Development Study)". In: *International Journal of Epidemiology* 35.1, pp. 34–41 (cit. on p. 6).
Quis, Johanna Sophie and Reif, Simon (2017). "Health Effects of Instruction Intensity Evidence from a Natural Experiment in German High-Schools Health EEects of Instruction Intensity Evidence from a Natural Experiment in German High-Schools We are grateful for helpful comments and suggestions". In: *FAU Discussion Papers in Economics* 12/2017 (cit. on p. 1).
Registrar General for England and Wales (1971). "1971 Census aggregate and Local Authority data. UK Data Service." In: (cit. on p. 12).
Richardson, Hannah (2016). *Grammar schools: What are they and why are they controversial?* (Cit. on pp. 2, 3).
Rubin, D (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies". In: *Journal of Educational Psychology* 66, pp. 688–701.
Rubin, Donald B and Thomas, Neal (1996). "Matching Using Estimated Propensity Scores: Relating Theory to Practice". In: *Biometrics* 52.1, pp. 249–264.
Silles, Mary A (2009). "Economics of Education Review The causal effect of education on health : Evidence from the United Kingdom". In: *Economics of Education Review* 28, pp. 122–128 (cit. on p. 1).
Stock, James H, Wright, Jonathan H, and Yogo, Motohiro (2002). "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments". In: *Journal of Business & Economic Statistics* 20.4, pp. 518–529 (cit. on p. 16).
Terza, Joseph V., Basu, Anirban, and Rathouz, Paul J. (2008). "Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling". In: *Journal of Health Economics* 27.3, pp. 531–543 (cit. on p. 13).
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, 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, 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, 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, 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, http://doi.org/10.5255/UKDA-SN-5565-.*
World Health Organisation (2017). *Body mass index - BMI* (cit. on p. 8).

9 Tables

Table 1: Descriptive statistics for all covariates by type of secondary school.

	Grammar				Comprehensive				Secondary modern			
	Mean	s.d.	Min	Max	Mean	s.d.	Min	Max	Mean	s.d.	Min	Max
Ability												
Cognitive skills age 7	0.76	0.10	0.35	1	0.61	0.16	0.04	0.99	0.59	0.15	0.09	0.94
Non-cognitive skills age 11	0.94	0.08	0.44	1	0.88	0.12	0.21	1	0.86	0.13	0.01	1
Relative cognitive ability age 11	0.79	0.15	0.04	1	0.50	0.29	0	1	0.37	0.22	0	1
General characteristics												
Female	0.55	0.50	0	1	0.48	0.50	0	1	0.49	0.50	0	1
Whether first born	0.36	0.48	0	1	0.31	0.46	0	1	0.30	0.46	0	1
Not white	0.02	0.14	0	1	0.04	0.20	0	1	0.04	0.20	0	1
Two or more siblings	0.65	0.48	0	1	0.73	0.45	0	1	0.75	0.43	0	1
Twin or triplet	0.01	0.11	0	1	0.02	0.15	0	1	0.03	0.17	0	1
Family background												
Mother's interest in child education	2.70	0.77	0	4	2.02	1.03	0	4	1.88	1.03	0	4
Father's SES high/middle-high	0.32	0.47	0	1	0.13	0.34	0	1	0.11	0.31	0	1
Father unemployed	0.01	0.09	0	1	0.03	0.17	0	1	0.04	0.18	0	1
Father job skilled/professional	0.54	0.50	0	1	0.47	0.50	0	1	0.47	0.50	0	1
No mother	0.00	0.06	0	1	0.01	0.08	0	1	0.01	0.09	0	1
No father	0.03	0.16	0	1	0.04	0.20	0	1	0.04	0.20	0	1
Council housing	0.19	0.39	0	1	0.39	0.49	0	1	0.39	0.49	0	1
Free school meals	0.03	0.16	0	1	0.09	0.29	0	1	0.10	0.30	0	1
Child in primary school												
Unhappy at school	0.03	0.18	0	1	0.07	0.26	0	1	0.07	0.26	0	1
Independent primary school	0.04	0.19	0	1	0.01	0.10	0	1	0.01	0.09	0	1
Child plans to study after school	0.43	0.50	0	1	0.23	0.42	0	1	0.17	0.37	0	1
Health endowment												
Maternal smoking during pregnancy	1.37	0.78	1	4	1.59	0.92	1	4	1.60	0.93	1	4

Child morbidity index	0.06	0.03	0	0	0.06	0.04	0	0	0.06	0.04	0	0
Chronic condition in the family	0.11	0.31	0	1	0.15	0.36	0	1	0.15	0.36	0	1
LEA characteristics in 1971												
Proportion comprehensive pupils in LEA	0.29	0.25	0	1	0.52	0.32	0	1	0.24	0.21	0	1
County level proportion unemp. male	0.04	0.02	0.02	0.10	0.04	0.02	0.02	0.10	0.04	0.02	0.02	0.10
— council housing	0.28	0.08	0.07	0.51	0.29	0.08	0.12	0.52	0.28	0.08	0.07	0.52
— owner-occupiers	0.49	0.16	0.01	0.76	0.48	0.14	0.01	0.70	0.52	0.11	0.01	0.76
— manufacturing employee	0.34	0.12	0.08	0.63	0.36	0.11	0.06	0.63	0.36	0.10	0.08	0.63
— agriculture employee	0.02	0.04	0	0.24	0.02	0.03	0	0.31	0.02	0.03	0	0.24
— lone parent families	0.09	0.02	0.06	0.16	0.10	0.02	0.06	0.16	0.09	0.02	0.06	0.16
— UK born men	0.91	0.06	0.78	0.98	0.91	0.06	0.78	0.99	0.92	0.05	0.78	0.98
— professional/managerial HOH	0.18	0.08	0.07	0.42	0.16	0.07	0.05	0.42	0.16	0.06	0.07	0.42
— non manual HOH	0.22	0.07	0.12	0.45	0.21	0.06	0.12	0.45	0.20	0.05	0.12	0.45
— skilled manual HOH	0.27	0.09	0.04	0.45	0.28	0.08	0.04	0.45	0.29	0.07	0.04	0.45
— semi-skilled manual HOH	0.11	0.04	0.01	0.21	0.12	0.04	0.01	0.21	0.12	0.03	0.01	0.21
— non-skilled manual HOH	0.07	0.02	0.01	0.13	0.07	0.02	0.03	0.17	0.07	0.02	0.01	0.14
County borough in 1971 census	0.26	0.44	0	1	0.34	0.47	0	1	0.27	0.44	0	1
London borough in 1971 census	0.11	0.31	0	1	0.10	0.29	0	1	0.04	0.20	0	1
Observations	1314				6135				2710			

Maternal smoking is measured on a 1-4 scale depending on intensity of smoking at the fourth month of pregnancy: no smoking, medium, variable, heavy.

Table 2: Descriptive statistics of health and wellbeing outcomes by type of secondary school attended

	Grammar				**Comprehensive**				**Secondary modern**			
	Mean	s.d.	Min	Max	Mean	s.d.	Min	Max	Mean	s.d.	Min	Max
Health outcomes												
Self-assessed health age 50	2.24	0.99	1	5	2.55	1.12	1	5	2.63	1.13	1	5
Excellent or very good SAH age 50	0.62	0.49	0	1	0.52	0.50	0	1	0.48	0.50	0	1
Malaise score age 50	1.27	1.73	0	9	1.52	1.97	0	9	1.59	2.00	0	9
Low malaise age 50	0.81	0.39	0	1	0.77	0.42	0	1	0.76	0.43	0	1
Mental ill-health score age 45	3.02	4.17	0	27	3.40	4.68	0	30	3.40	4.63	0	30
BMI measured age 45	26.41	4.61	17	51	27.56	4.88	17	54	27.67	5.16	18	64
Cholesterol ratio age 45	3.80	1.15	2	8	3.97	1.17	2	10	4.07	1.18	2	12
Triglyceride age 45	1.88	1.46	0	17	2.06	1.61	0	25	2.15	1.71	0	27
Fibrinogen g/L age 45	2.88	0.56	1	5	2.98	0.63	1	7	3.00	0.62	1	6
C reactive protein g/L age 45	1.84	3.35	0	34	2.27	4.93	0	152	2.26	4.26	0	94
Well-being outcomes												
School aspirations age 16 (PCA)	1.12	1.31	-3	2	-0.17	1.50	-4	2	-0.48	1.32	-4	2
Work aspirations age 16 (dummy)	0.93	0.26	0	1	0.79	0.41	0	1	0.76	0.43	0	1
Life satisfaction age 33 (PCA)	-0.04	1.40	-8	2	0.04	1.44	-8	2	-0.02	1.49	-8	2
Self-efficacy age 33 (PCA)	0.21	1.17	-5	1	-0.02	1.34	-5	1	-0.06	1.34	-5	1
Positive about job age 33 (PCA)	0.37	1.20	-4	2	-0.03	1.41	-5	2	-0.15	1.44	-5	2
Contact with police age 45	0.14	0.35	0	1	0.18	0.38	0	1	0.18	0.38	0	1
Ever tried illegal drugs age 45	0.19	0.39	0	1	0.17	0.38	0	1	0.17	0.37	0	1
Observations	1308				6002				2651			

Note that the self-assessed health scale goes from 1)excellent health to 5)very poor health.

Table 3: Pre- and post-matching moments of key covariates for the GC sample.

	Grammar			**Comprehensive**		
	Mean	Variance	Skewness	Mean	Variance	Skewness
Raw sample						
Cognitive skills	0.763	0.010	-0.467	0.618	0.025	-0.404
Non-cognitive skills	0.940	0.006	-2.288	0.882	0.015	-1.513
Relative cognitive score	0.795	0.021	-0.944	0.507	0.082	-0.029
Mother's interest in edu	2.697	0.585	-1.843	2.027	1.057	-0.490
High father's SES dummy	0.822	0.146	-1.685	0.692	0.213	-0.831
After						
Cognitive skills	0.763	0.010	-0.467	0.763	0.010	-0.469
Non-cognitive skills	0.940	0.006	-2.288	0.940	0.006	-2.286
Relative cognitive score	0.795	0.021	-0.944	0.795	0.021	-0.946
Mother's interest in edu	2.697	0.585	-1.843	2.697	0.585	-1.842
High father's SES dummy	0.822	0.146	-1.685	0.822	0.146	-1.683

Table 4: Pre- and post-matching moments of key covariates for the SMC sample.

	Secondary modern			**Comprehensive**		
	Mean	Variance	Skewness	Mean	Variance	Skewness
Raw sample						
Cognitive skills	0.590	0.022	-0.308	0.618	0.025	-0.404
Non-cognitive skills	0.867	0.016	-1.363	0.882	0.015	-1.513
Relative cognitive score	0.376	0.047	0.363	0.507	0.082	-0.029
Mother's interest in edu	1.908	1.065	-0.317	2.027	1.057	-0.490
High father's SES dummy	0.671	0.221	-0.728	0.692	0.213	-0.831
After						
Cognitive skills	0.590	0.022	-0.308	0.590	0.022	-0.308
Non-cognitive skills	0.867	0.016	-1.363	0.867	0.016	-1.362
Relative cognitive score	0.376	0.047	0.363	0.376	0.047	0.364
Mother's interest in edu	1.908	1.065	-0.317	1.908	1.065	-0.317
High father's SES dummy	0.671	0.221	-0.728	0.671	0.221	-0.728

Table 5: Models for health outcomes (matched grammar and comprehensive sample).

	High SAH	**Low malaise**	**Mental ill-health**	**BMI**	**Chol ratio**	**Triglycerides**	**CRP**	**Fibrinogen**
Grammar	-0.0050	0.0662	0.0759	-0.5154+	0.0457	-0.0109	0.0133	0.0086
	(0.0287)	(0.0904)	(0.2411)	(0.2819)	(0.0677)	(0.0892)	(0.2484)	(0.0378)
Cognitive skills	0.2214	0.2750	-0.7180	-2.9662*	-0.6453+	-0.9102*	-0.5829	-0.0519
	(0.1497)	(0.4695)	(1.2463)	(1.4501)	(0.3479)	(0.4572)	(1.2672)	(0.1927)
Non-cognitive skills	0.2276	0.3679	-2.8256+	-2.0607	-0.3988	-0.3938	-1.0304	-0.3193
	(0.1908)	(0.6078)	(1.5962)	(1.8650)	(0.4492)	(0.5922)	(1.6475)	(0.2506)
Relative cogn. ability	0.0896	0.4099	1.2666	0.1419	-0.2651	-0.1241	-2.1491*	-0.2749+
	(0.1081)	(0.3325)	(0.9091)	(1.0628)	(0.2580)	(0.3393)	(0.9462)	(0.1443)
Female	0.0155	-0.3741***	1.0580***	-1.1305***	-0.9983***	-1.0672***	0.0497	0.1372***
	(0.0292)	(0.0928)	(0.2433)	(0.2840)	(0.0687)	(0.0906)	(0.2520)	(0.0383)
Mother's interest	0.0066	0.0293	0.0869	-0.3400+	-0.0303	-0.0882	-0.0973	-0.0409
	(0.0199)	(0.0622)	(0.1727)	(0.2028)	(0.0495)	(0.0651)	(0.1816)	(0.0276)
Father's SES	0.0125	-0.2266	-0.0799	-0.4781	-0.0164	-0.1864	-0.0769	-0.0238
	(0.0430)	(0.1419)	(0.3543)	(0.4143)	(0.1027)	(0.1352)	(0.3777)	(0.0574)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2875	2854	1141	1135	951	954	940	937
F			1.4594	2.0902	7.2168	5.6455	0.5411	1.5607
chi2	52.8400	56.4290						

Standard errors in parentheses. Mother's interest in child education is measured on a scale 0-4, and father's SES is a dummy for high/middle-high SES.
+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table 6: Models for wellbeing outcomes (matched grammar and comprehensive sample).

	School	**Work**	**Life sat.**	**Self-eff.**	**Job positiv.**	**Crime**	**Drugs**
Grammar	0.3065***	0.0130	-0.1882*	-0.0710	0.0010	-0.0149	-0.0098
	(0.0656)	(0.0141)	(0.0808)	(0.0654)	(0.0665)	(0.0202)	(0.0213)
Cognitive skills	0.2472	-0.0614	-0.0857	0.7847*	0.3044	0.0767	-0.0107
	(0.3420)	(0.0733)	(0.4260)	(0.3449)	(0.3502)	(0.1056)	(0.1107)
Non-cognitive skills	1.1752**	0.0608	1.7151**	1.4856***	0.0866	-0.0491	-0.3042*
	(0.4295)	(0.0900)	(0.5491)	(0.4440)	(0.4496)	(0.1265)	(0.1294)
Relative cogn. ability	2.2181***	0.1842***	-0.1687	0.4725+	1.1453***	-0.0725	0.1405+
	(0.2433)	(0.0498)	(0.3034)	(0.2455)	(0.2496)	(0.0751)	(0.0814)
Female	-0.0007	0.0107	0.0904	-0.0927	-0.5468***	-0.1075***	-0.0663**
	(0.0662)	(0.0142)	(0.0819)	(0.0663)	(0.0673)	(0.0202)	(0.0213)
Mother's interest	0.1632***	0.0019	0.0821	-0.0147	0.0928*	0.0227	0.0210
	(0.0439)	(0.0091)	(0.0553)	(0.0448)	(0.0460)	(0.0146)	(0.0151)
Father's SES	0.0411	0.0190	0.1180	-0.0490	0.0783	-0.0780**	-0.0072
	(0.0976)	(0.0196)	(0.1207)	(0.0977)	(0.0996)	(0.0280)	(0.0315)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1599	4156	1245	1229	1258	3277	3279
F	8.1650		1.2206	1.7262	4.1799		

Standard errors in parentheses. Mother's interest in child education is measured on a scale 0-4, and father's SES is a dummy for high/middle-high SES.
+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table 7: Models for health outcomes (matched secondary modern and comprehensive sample).

	High SAH	**Low malaise**	**Mental ill-health**	**BMI**	**Chol ratio**	**Triglycerides**	**CRP**	**Fibrinogen**
Secondary modern	-0.0017	0.0247	-0.1593	0.1753	0.0802	0.0415	-0.1227	-0.0177
	(0.0232)	(0.0673)	(0.2213)	(0.2405)	(0.0583)	(0.0819)	(0.2732)	(0.0324)
Cognitive skills	0.0507	0.0986	-0.6295	-0.4228	-0.0392	-0.1484	-1.1132	-0.0293
	(0.1007)	(0.2905)	(0.9556)	(1.0403)	(0.2531)	(0.3566)	(1.1814)	(0.1401)
Non-cognitive skills	0.3476***	1.1693***	-2.8631**	0.5798	-0.3138	-0.3525	0.8866	0.0563
	(0.0964)	(0.2708)	(0.9431)	(1.0205)	(0.2522)	(0.3544)	(1.1566)	(0.1372)
Relative cogn. ability	0.1510*	0.4131*	-0.6995	-0.6991	-0.1177	-0.1347	-0.7343	-0.2764**
	(0.0700)	(0.2064)	(0.6673)	(0.7258)	(0.1774)	(0.2500)	(0.8292)	(0.0983)
Female	-0.0362	-0.3900***	1.2586***	-0.8255***	-0.7761***	-0.8864***	0.5570*	0.1685***
	(0.0228)	(0.0674)	(0.2180)	(0.2373)	(0.0578)	(0.0812)	(0.2708)	(0.0321)
Mother's interest	0.0269*	-0.0557+	0.0752	-0.3877**	-0.0578*	-0.0112	-0.0557	-0.0235
	(0.0114)	(0.0335)	(0.1108)	(0.1204)	(0.0292)	(0.0411)	(0.1370)	(0.0162)
Father's SES	0.0234	-0.0127	0.2579	-0.7486**	-0.0043	-0.0362	-0.1394	-0.0073
	(0.0276)	(0.0799)	(0.2603)	(0.2837)	(0.0691)	(0.0972)	(0.3233)	(0.0383)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3250	3224	1878	1853	1579	1581	1553	1550
F			3.0529	2.3501	6.3015	4.1947	1.8252	2.4136
chi2	109.2700	96.6975						

Standard errors in parentheses. Mother's interest in child education is measured on a scale 0-4, and father's SES is a dummy for high/middle-high SES.

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

Table 8: Models for wellbeing outcomes (matched secondary modern and comprehensive sample).

	School	**Work**	**Life sat.**	**Self-eff.**	**Job positiv.**	**Crime**	**Drugs**
Secondary modern	0.1483**	0.0176	-0.0093	0.1075+	0.0377	-0.0221	-0.0064
	(0.0474)	(0.0155)	(0.0659)	(0.0580)	(0.0604)	(0.0161)	(0.0155)
Cognitive skills	-0.1316	0.0339	0.2809	0.4825+	0.3378	0.0883	0.1056
	(0.1974)	(0.0633)	(0.2817)	(0.2471)	(0.2582)	(0.0688)	(0.0659)
Non-cognitive skills	0.8770***	0.1315*	0.8366**	0.5058*	0.6298*	-0.1287*	-0.2384***
	(0.2010)	(0.0630)	(0.2825)	(0.2484)	(0.2598)	(0.0643)	(0.0609)
Relative cogn. ability	1.9199***	0.4149***	-0.2897	0.2951+	0.8434***	-0.0409	0.0326
	(0.1415)	(0.0477)	(0.1987)	(0.1741)	(0.1820)	(0.0486)	(0.0464)
Female	0.0214	-0.0052	0.2207***	-0.0316	-0.8009***	-0.1968***	-0.0832***
	(0.0467)	(0.0153)	(0.0646)	(0.0567)	(0.0592)	(0.0155)	(0.0152)
Mother's interest	0.1610***	0.0141+	0.0130	0.0306	0.0380	0.0055	0.0095
	(0.0238)	(0.0077)	(0.0330)	(0.0290)	(0.0303)	(0.0081)	(0.0078)
Father's SES	0.1633**	0.0311+	0.0646	0.1037	0.0700	-0.0128	-0.0215
	(0.0556)	(0.0177)	(0.0775)	(0.0681)	(0.0712)	(0.0190)	(0.0180)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2766	4818	2127	2105	2135	3777	3779
F	21.0253		2.1487	3.0606	9.4123		

Standard errors in parentheses. Mother's interest in child education is measured on a scale 0-4, and father's SES is a dummy for high/middle-high SES.
+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table 9: First stage for grammar and secondary modern attendance, using % pupils going to comprehensive schools in individual's LEA as an IV

	(1)	(2)	(3)	(4)
	Grammar	Grammar	Sec modern	Sec modern
% comprehensive pupils in LEA	-0.6193***	-0.7216***	-0.7546***	-0.7495***
	(0.0339)	(0.0381)		(0.0230)
(0.0266)				
Cognitive ability		0.0466		-0.0020
		(0.1175)		(0.0674)
Non-cognitive skills		-0.0784		0.0449
		(0.1466)		(0.0675)
Relative cogn. abi.		-0.1151		-0.0536
		(0.0827)		(0.0482)
Observations	5467	4412	6396	4807
F	166.9369	13.1494	1072.4188	25.3480

Standard errors in parentheses

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

Table 10: IV models for health outcomes (matched grammar and comprehensive sample).

	High SAH	**Low malaise**	**Mental ill-health**	**BMI**	**Chol ratio**	**Triglycerides**	**CRP**	**Fibrinogen**
Grammar	0.0073	0.0802	0.1815	-1.1399+	-0.0011	-0.1513	-0.3904	-0.0224
	(0.0663)	(0.0534)	(0.5495)	(0.6548)	(0.1579)	(0.2087)	(0.5827)	(0.0883)
Cognitive skills	0.2186	0.0720	-0.7230	-2.9404*	-0.6436+	-0.9076*	-0.5926	-0.0527
	(0.1527)	(0.1224)	(1.2468)	(1.4537)	(0.3482)	(0.4579)	(1.2696)	(0.1928)
Non-cognitive skills	0.2312	0.1008	-2.8320+	-2.0508	-0.3974	-0.3908	-1.0253	-0.3181
	(0.1966)	(0.1587)	(1.5969)	(1.8694)	(0.4496)	(0.5931)	(1.6506)	(0.2508)
Relative cogn. ability	0.0913	0.1212	1.2837	0.0548	-0.2688	-0.1331	-2.1683*	-0.2765+
	(0.1109)	(0.0892)	(0.9128)	(1.0685)	(0.2585)	(0.3400)	(0.9483)	(0.1445)
Female	0.0136	-0.0988***	1.0542***	-1.1157***	-0.9969***	-1.0631***	0.0571	0.1378***
	(0.0298)	(0.0240)	(0.2440)	(0.2850)	(0.0689)	(0.0909)	(0.2527)	(0.0384)
Mother's interest	0.0061	0.0080	0.0883	-0.3504+	-0.0311	-0.0908	-0.1012	-0.0412
	(0.0206)	(0.0166)	(0.1729)	(0.2035)	(0.0496)	(0.0653)	(0.1820)	(0.0276)
Father's SES	0.0129	-0.0517	-0.0765	-0.5026	-0.0179	-0.1914	-0.0954	-0.0251
	(0.0439)	(0.0353)	(0.3547)	(0.4159)	(0.1028)	(0.1356)	(0.3791)	(0.0575)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2875	2854	1141	1135	951	954	940	937
F	1.3807	1.5620	1.4590	2.0726	7.1932	5.6424	0.5509	1.5584

Standard errors in parentheses. Mother's interest in child education is measured on a scale 0-4, and father's SES is a dummy for high/middle-high SES.

+ $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 11: IV models for wellbeing outcomes (matched grammar and comprehensive sample).

	School	**Work**	**Life sat.**	**Self-eff.**	**Job positiv.**	**Crime**	**Drugs**
Grammar	0.1689	0.0027	-0.2341	-0.0465	-0.0203	-0.0325	0.0726
	(0.1549)	(0.0337)	(0.1873)	(0.1514)	(0.1546)	(0.0467)	(0.0493)
Cognitive skills	0.2396	-0.0776	-0.0804	0.7816*	0.3067	0.0802	-0.0256
	(0.3426)	(0.0749)	(0.4267)	(0.3454)	(0.3506)	(0.1072)	(0.1131)
Non-cognitive skills	1.1666**	0.0743	1.7165**	1.4853***	0.0865	-0.0532	-0.3297*
	(0.4302)	(0.0936)	(0.5494)	(0.4440)	(0.4496)	(0.1342)	(0.1416)
Relative cogn. ability	2.2040***	0.2050***	-0.1737	0.4748+	1.1431***	-0.0740	0.1487+
	(0.2440)	(0.0534)	(0.3041)	(0.2459)	(0.2501)	(0.0776)	(0.0819)
Female	0.0048	0.0123	0.0923	-0.0937	-0.5460***	-0.1098***	-0.0694**
	(0.0665)	(0.0145)	(0.0823)	(0.0665)	(0.0675)	(0.0209)	(0.0221)
Mother's interest	0.1610***	0.0015	0.0816	-0.0143	0.0926*	0.0234	0.0221
	(0.0440)	(0.0096)	(0.0554)	(0.0449)	(0.0461)	(0.0144)	(0.0152)
Father's SES	0.0389	0.0228	0.1166	-0.0482	0.0774	-0.0848**	-0.0048
	(0.0977)	(0.0213)	(0.1208)	(0.0978)	(0.0998)	(0.0303)	(0.0320)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1599	1617	1245	1229	1258	1287	1288
F	7.6007	1.3948	1.1180	1.6972	4.1800	1.7519	1.7779

Standard errors in parentheses. Mother's interest in child education is measured on a scale 0-4, and father's SES is a dummy for high/middle-high SES.
+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table 12: IV models for health outcomes (matched secondary modern and comprehensive sample).

	High SAH	**Low malaise**	**Mental ill-health**	**BMI**	**Chol ratio**	**Triglycerides**	**CRP**	**Fibrinogen**
Secondary modern	-0.0024	-0.0265	-0.2538	-0.1115	0.1580	0.1505	-0.0386	-0.0123
	(0.0538)	(0.0462)	(0.5052)	(0.5542)	(0.1328)	(0.1872)	(0.6271)	(0.0742)
Cognitive skills	0.0497	0.0287	-0.6293	-0.4164	-0.0385	-0.1482	-1.1131	-0.0292
	(0.1016)	(0.0875)	(0.9557)	(1.0410)	(0.2533)	(0.3569)	(1.1815)	(0.1401)
Non-cognitive skills	0.3466***	0.3682***	-2.8634**	0.5765	-0.3121	-0.3512	0.8893	0.0565
	(0.0973)	(0.0838)	(0.9432)	(1.0211)	(0.2524)	(0.3547)	(1.1568)	(0.1372)
Relative cogn. ability	0.1529*	0.1240*	-0.7003	-0.7024	-0.1172	-0.1342	-0.7323	-0.2763**
	(0.0711)	(0.0612)	(0.6674)	(0.7262)	(0.1776)	(0.2502)	(0.8294)	(0.0984)
Female	-0.0362	-0.1133***	1.2606***	-0.8201***	-0.7777***	-0.8886***	0.5552*	0.1684***
	(0.0231)	(0.0199)	(0.2182)	(0.2376)	(0.0579)	(0.0813)	(0.2711)	(0.0322)
Mother's interest	0.0271*	-0.0168+	0.0744	-0.3906**	-0.0573+	-0.0105	-0.0551	-0.0235
	(0.0116)	(0.0100)	(0.1108)	(0.1206)	(0.0293)	(0.0411)	(0.1371)	(0.0163)
Father's SES	0.0225	-0.0070	0.2575	-0.7478**	-0.0046	-0.0368	-0.1385	-0.0073
	(0.0279)	(0.0241)	(0.2603)	(0.2839)	(0.0691)	(0.0972)	(0.3233)	(0.0383)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Observations	3250	3224	1878	1853	1579	1581	1553	1550
F	2.9665	2.6751	3.0451	2.3339	6.2781	4.1983	1.8196	2.4052

Standard errors in parentheses. Mother's interest in child education is measured on a scale 0-4, and father's SES is a dummy for high/middle-high SES.

+ $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 13: IV models for wellbeing outcomes (matched secondary modern and comprehensive sample).

	School	**Work**	**Life sat.**	**Self-eff.**	**Job positiv.**	**Crime**	**Drugs**
Secondary modern	0.1858+	0.0531	-0.0009	0.0867	-0.0561	-0.0270	0.0074
	(0.1023)	(0.0342)	(0.1446)	(0.1275)	(0.1341)	(0.0372)	(0.0355)
Cognitive skills	-0.1309	0.0666	0.2812	0.4812+	0.3313	0.0952	0.1075
	(0.1975)	(0.0655)	(0.2818)	(0.2473)	(0.2585)	(0.0701)	(0.0668)
Non-cognitive skills	0.8770***	0.1616*	0.8364**	0.5062*	0.6317*	-0.1420*	-0.2604***
	(0.2010)	(0.0662)	(0.2825)	(0.2485)	(0.2599)	(0.0689)	(0.0656)
Relative cogn. ability	1.9223***	0.3926***	-0.2896	0.2955+	0.8432***	-0.0525	0.0345
	(0.1416)	(0.0470)	(0.1987)	(0.1741)	(0.1821)	(0.0495)	(0.0472)
Female	0.0211	-0.0059	0.2205***	-0.0312	-0.7990***	-0.2005***	-0.0830***
	(0.0467)	(0.0154)	(0.0647)	(0.0567)	(0.0593)	(0.0161)	(0.0154)
Mother's interest	0.1611***	0.0147+	0.0131	0.0304	0.0372	0.0052	0.0097
	(0.0238)	(0.0079)	(0.0331)	(0.0290)	(0.0303)	(0.0082)	(0.0078)
Father's SES	0.1639**	0.0349+	0.0647	0.1033	0.0686	-0.0139	-0.0217
	(0.0556)	(0.0183)	(0.0776)	(0.0682)	(0.0713)	(0.0194)	(0.0185)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2766	2870	2127	2105	2135	2246	2247
F	20.8419	7.7455	2.1480	2.9785	9.3945	5.7980	2.3666

Standard errors in parentheses. Mother's interest in child education is measured on a scale 0-4, and father's SES is a dummy for high/middle-high SES.
+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001