UNIVERSITY of York

HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

WP 18/32

Tracking pupils into adulthood: selective schools and long-term well-being in the 1958 British cohort

Andrew M. Jones; Chiara Pastore and Nigel Rice

November 2018

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

Tracking pupils into adulthood: selective schools and long-term well-being in the 1958 British cohort

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

Abstract

We explore the effect of tracking pupils by ability into different secondary schools on adult health, well-being and labour outcomes in England. We address selection bias by balancing individual pre-treatment characteristics via entropy matching, followed by parametric regressions estimated via OLS and IV approaches. Ability tracking does not affect long-term health and well-being, while it marginally raises hourly wages for low-ability pupils, compared to a mixed-ability system. Cognitive and non-cognitive abilities measured prior to secondary school are more significant and positive predictors of adult outcomes. Particularly, non-cognitive skills may have a protective role for adult health for lower cognitive ability children.

Keywords Ability tracking, Educational reform, Well-being, Health, Entropy balancing, Instrumental variables.

JEL I26, I28, I1, 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

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 Thomas Cornelissen, as well as participants at the HEDG seminars at the University of York, the NCDS 60 Years of Our Lives conference, the RES Junior Symposium 2018 at the University of Sussex, the 4th IRDES-Dauphine AHEPE Workshop and the 3rd IZA Workshop: The Economics of Education for helpful comments.

1 Introduction

Educational policy may be one of the most effective tools to improve life opportunities for individuals across all backgrounds. A central issue in the provision of public education is whether and how to tailor the curriculum around pupils' ability. Tracking students by ability into different schools can be seen as a way to improve learning and teaching efficiency, while reducing socioeconomic inequalities, since selection into prestigious institutions is based on academic talent, in principle regardless of family circumstances[1]. A counter-argument is that such systems are naturally biased to favour children from affluent backgrounds, since these pupils are generally more prepared to take entry tests and are more supported by their families in their schooling choices. If this is true, selecting students by ability at a young age would have detrimental effects on the pre-existing inequality gap.

Research on the consequences of school tracking often focuses on education and labour outcomes. However, the non-monetary benefits of education are also broad, and accrue over time through a variety of pathways, including health and well-being (Grossman, 1972). Based on the idea that early investments in human capital boost self-productivity, more and better education is likely to improve children's cognitive and non-cognitive skills, shaping their preferences over time and risk, determining their future peer networks and habit formation, thus leading to better health and higher life satisfaction (Campbell et al., 2014, Cutler and Lleras-Muney, 2011, Fuchs, 1982). Assessing the role of specific educational policies as early determinants of health and well-being can, on the one hand, help national preventive strategies, urgently needed by healthcare and welfare systems pressured by ever-increasing costs around the world. On the other, it can contribute to explaining persisting income inequalities, given that good health and general well-being are requirements for a successful and productive work life.

We assess the long-term health and well-being effects of attending secondary school in a system where an ability test determines entry into a prestigious institution, compared to attending secondary school in a non-selective system, where institutions receive pupils of all abilities. 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 high-quality information on their education, health and personal history collected over time. We exploit the comprehensive schooling reform implemented in England and Wales in the 1960s, a time in which some NCDS children were exposed to a selective secondary schooling system, with selective grammar schools offering the more academic track, and secondary modern, more vocational, being the main alternative. The remaining NCDS cohort attended school in a

[1] A different policy, not covered here, is tracking student into different classes, but within the same school (see Burgess, 2016).

comprehensive system, where school assignment was not linked to ability. The system experienced largely depended on Local Education Authority (LEA) of residence. Attendance to different school types exposed pupils to different curricula, teacher quality and peer ability, thus offering an opportunity to explore the effect of variation in quality. Table 1 summarises school characteristics by type for NCDS participants. Grammar schools displayed the lowest pupil teacher ratio of all, and the most pupils passing 2 or more A-levels and doing a degree when leaving school. Higher proportions of grammar schools were single sex and had sports facilities, while secondary modern schools were more likely to lack science labs and sports facilities than other schools.

Our project is timely in providing evidence to inform the current policy debate on the reintroduction of selective schools in the UK, where the government has recently pledged a £50 million investment towards new selective schools. Several other countries incorporate selection in their secondary schooling systems, including Australia, France, Germany, the Netherlands, Switzerland and the United States, and research on the effects of school tracking by ability on education and labour outcomes is abundant (see Burgess, 2016 for a review). Selection is generally associated with lower social mobility, and higher inequality in income and educational attainment (Brunello and Checchi, 2007; Burgess, Dickson, et al., 2014; Hanushek and Wößmann, 2006). The NCDS study has been used in the past to evaluate selective schooling policy in the UK in terms of its impact on education and labour outcomes (Galindo-Rueda and Vignoles, 2005; Kerckhoff, 1986, 1996). 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). A more recent study on the long-term impact of elite schools on education, labour, marriage and fertility for the marginal student focused on a Scottish cohort and had no specific health measure (Del Bono and Clark, 2016).

In our analysis we look at two parallel questions. On the one hand, we explore long-term effects of attending grammar, compared to comprehensive, to help answer the question about the impact of reintroducing selective schools today. On the other, we investigate the effect of attending secondary modern, compared to comprehensive. Today's mixed-ability schools found in the same areas as new selective schools would likely experience similar effects to those experienced by secondary modern schools in the past, even if only in terms of pupil and resource allocation[2]. Separating treatment effects allows us to make treatment and control groups more comparable, but we additionally implement entropy balancing as a way to balance the data, to achieve higher confidence that we are able to estimate an unbiased treatment effect. The balancing exercise is followed by parametric regressions for a rich set of outcomes. With a similar approach to that

[2]Overall the share of school leavers going to University today is significantly higher, and school type often determines the rank of the University, rather than whether the pupil gets a degree, and therefore a professional, manual or other type of occupation.

implemented by Manning and Pischke (2006), we are able to show that our analysis allows us to overcome selection issues noted previously in the literature, 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). However, our study is unique in linking types of school, 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, with the exception of labour ones, while childhood cognitive and non-cognitive abilities have a long-lasting association with several adult outcomes.

Section 2 briefly reviews the history of selective schooling in the English education system, and the existing knowledge on its effects. Sections 3 and 4 outline the data and a reference framework for the relationship between schooling and later outcomes. We then describe our two-step approach to estimate the treatment effects of interest, which addresses the main challenges of this literature. Entropy balancing is combined with parametric regressions, 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 Selection by ability in England

The origins of tracking in the British secondary schooling system go back to the 1944 Education Act, which established the reorganisation of state secondary schools by 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). The 11-plus was set at LEA level, so difficulty and entry score varied across the country.

Given the growing dissatisfaction with the allocation system in state schools, the Labour government started phasing out the selective system of secondary schooling with Circular 10/65 in 1965. 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 (Bolton, 2017; Galindo-Rueda and Vignoles, 2005). 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). At the time of writing, 163 grammar schools exist in England, attended by approximately 167,000 pupils (Richardson, 2016). New talks about the reintroduction of selective schools started in 2016, when the newly appointed Prime Minister Theresa May pledged support for new selective schools as a means towards a meritocratic education system. The most recent announcement concerns a £50 million investment for grammar schools (Long et al., 2017), thus opening space for research seeking evidence on the potential impact of a more selective education system for present and future generations of pupils.

The 1960s comprehensive reform in England offers an opportunity to evaluate long-term effects of selection by ability in secondary schooling. 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, due to unobservable factors that determine both education and outcomes. In the present case, higher well-being in adulthood and type of school attended could both be influenced by individual ability or parental investments. The literature has dealt with this issue in different ways, mainly in the estimation of effects on earnings and educational achievement. Using NCDS data and an instrumental variable strategy, Galindo-Rueda and Vignoles (2005) instrument comprehensive school attendance with political control in the individual's electoral constituency and share of comprehensive schools in the individual's LEA. Their results suggest that the shift to comprehensive schooling reduced educational achievement for more able children only. The validity of this type of analysis was put under scrutiny by Manning and Pischke (2006). They criticise value-added approaches comparing outcomes for pupils in selective and comprehensive areas that only added pre-secondary school outcomes as controls, arguing they are not sufficient to remove the selection problem arising from fundamental differences between the two groups. This is shown by running a placebo regression of pre-secondary school test scores on an indicator for comprehensive school attendance. The treatment effect is significant in all of their specifications, including the ones using political control of the county as instrument, following Galindo-Rueda and Vignoles (2005), and this is taken as evidence against this type of empirical strategy. 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.

A number of studies have used alternative methods that are more robust to the criticisms advanced above. Using a nationally representative British household panel, Burgess, Dickson, 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 for pupils who attended school in selective areas. Burgess, Crawford, et al. (2017) further 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 areas pupils with similar school exam scores however, grammar students did not do significantly better at University (Burgess, Crawford, et al., 2017). Using Scottish data, Del Bono and Clark (2016) 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 the 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, and Rosa Dias (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 percentage 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. Our analysis builds on this previous work and makes two important contributions. The entropy balancing algorithm increases comparability, supporting the credibility of our strategy in the face of the selection problem. Second, the range of outcomes we consider allows us to build a well-rounded picture of non-monetary returns of selective versus non-selective secondary school at different points of the individual's lives. The hope is that this broad scope can help us understand more about the paths that lead from education to adult inequalities in health, income and general well-being.

[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.

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 at birth with a sample of over 17,000 individuals, and retained about 9,000 at the most recent wave in 2013 (Brown et al., 2016)[4]. 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[5]. The key variables for the present study are described below.

3.1 Pre-treatment characteristics

Detailed information from the first three waves of the survey allows us to control for a broad set of pre-secondary 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). Family background covariates include mother's interest in child's education (expressed on a 0-4 scale), father's employment status and socioeconomic status (SES), family composition, financial hardship and council housing during childhood. Rich information is available on infant and child health, which is likely to affect both schooling and long-term health outcomes. We group childhood health conditions from twelve categories under one single indicator of child morbidity, following previous literature (Jones, Rice, and Rosa Dias, 2011; Power and Elliott, 2006). Maternal smoking during pregnancy, presence of chronic conditions in the family, and hospital admissions up to age 7 are included to reflect health endowment. Data collected at age 7 and 11 also includes information on primary school. Finally, local area characteristics, based on LEA of school attended in 1974, were retrieved from the 1971 Census.

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. 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

[4]During childhood, participants were traced through the school system, while in adulthood tracing them became more difficult, which is one of the main reason for attrition over time (Power and Elliott, 2006).

[5]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).

(ESN), and other special needs). In this paper, we consider only the first three categories, 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). 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).

3.3 Ability

Cognitive skills were assessed through numeracy, reading, verbal and non-verbal tests at ages 7, 11 and 16. That is, during primary, just before secondary and again just after secondary school respectively. Following existing literature, we grouped test scores under a single indicator of cognitive ability for each age, by implementing principal component analysis (PCA) (Cawley et al., 1996; Galindo-Rueda and Vignoles, 2005; Jones, Rice, and Rosa Dias, 2011). PCA captures the variation in the data, while avoiding multi-collinearity issues that would arise if all the test scores were included as regressors in the model. As noted by Jones, Rice, and Rosa Dias (2011), age 11 tests closely resemble the three components of the 11-plus: mathematics, reading, verbal and non-verbal ability. Performing PCA, 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 Appendix A.1). 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. For simplicity of interpretation, we then converted the PCA indices to variables bound between 0 and 1. Further, a rank variable is constructed from the age 11 cognitive ability index, ranking individuals by their measured cognitive ability. This was 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 11. Teachers were asked to answer questions on twelve child behaviour dimensions. The score was converted to a variable bound between 0 and 1, so that it is increasing in non-cognitive skills. Due to the way the questionnaire was designed, its distribution is highly skewed towards the right (no behavioural problems).

3.4 Outcomes

3.4.1 Well-being and labour measures

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. 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. 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 twelve types of drugs, or any other illegal drug, except cannabis. We also examine two labour outcomes both at ages 33 and 50. The first is individual gross hourly wage, imputed from weekly, monthly or bi-monthly income and hours worked per week, and then log-transformed for regression analysis. The second is a dummy indicating whether the individual is in employment at the time.

3.4.2 Survey health measures

Long-term impact of treatment is also assessed on a broad range of health dimensions. 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 questionnaire offers a measure of ill-health and discomfort, both physical and mental. For ease of interpretation, a binary variable for excellent or good SAH and a binary variable for low malaise (scores 0,1 or 2) are used as self-assessed health outcomes at age 50. 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.

3.4.3 Biometric health measures

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 $BMI > 30$. High BMI 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. Our outcomes include C-Reactive protein (CRP) (g/L), fibrinogen (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. The use of biomedical outcomes represents an original element of our study in the literature on the effects of school quality, as it allows us to characterise health status on a finer scale. In addition to current health problems, we are thus able to assess the effect of education on the risk of presenting health problems in the future.

3.4.4 Attrition

As in most longitudinal studies, a concern when analysing the data is that attrition can be non-random. Specifically, if attrition is correlated with variables related to the treatment or outcome of interest, then estimates of treatment effect could be biased. For each survey wave that we use, we therefore examine some key cohort members' characteristics measured at birth and in childhood, as well percentages attending each type of secondary school. Table A4 in the Appendix shows that there are small but noticeable differences between average characteristics in the sample of people dropped from all analyses because of attrition or missing data, and the sample used for outcome analysis at each wave. Dropped individuals are less likely to be first born and their mother is more likely to have left school before legal school-leaving age and to have smoked more frequently during pregnancy. Note however that treatment status, school type, is observed at 16. Since all samples used for the analysis from age 16 onwards present hardly any differences in average characteristics shown, this increases confidence that sample composition does not vary systematically in relation to key characteristics after this point in time, and this is particularly reassuring for our analysis, in agreement with other literature (Case et al., 2005; Jones, Rice, and Rosa Dias, 2011). An important feature is that the percentage of pupils attending each type of school does not vary over time. In any case, as noted by Dearden et al. (2002), even if lower ability and lower SES pupils were under-represented in the sample, the fact that we control for such characteristics in our analysis minimises the possibility of bias in treatment effect estimates[6].

4 Methods

For our model we draw from the framework of individual investment in own human capital, represented here by health and other well-being dimensions. We assume three time periods t =*0,1,2*, corresponding 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, and

[6]Survey weights provided by the Centre for Longitudinal Studies would not be of help in this context because they are not aimed at balancing school type over time.

with a genetic endowment of cognitive and non-cognitive abilities A_{i0}. These characteristics 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, SU_{i1}.

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

The production functions for adult health and well-being outcomes, Y_{i2}, depend on background, pre-secondary school ability, type of school (our treatment variable), and local area characteristics[7].

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

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 the equation for 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 our context. 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 (Bonhomme and Sauder, 2011; Manning and Pischke, 2006).

Isolating treatment effects to establish more than simple correlations requires comparing treated individuals with credible counterfactuals (Heckman et al., 1997; D. Rubin, 1974). 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:

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

Similarly, we estimate the effect of going to secondary modern, compared to comprehensive, by contrasting secondary modern pupils and comprehensive pupils who would have

[7] 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.

attended secondary modern, had they experienced the selective system[8]:

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

The way we ensure we compare like with like is via entropy balancing, 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 balancing procedure[9]. The regressions 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[10]. Then, balancing covariates for treatment and control groups and using resulting weights in subsequent parametric approaches can help reduce model dependence on crucial, although not entirely verifiable, parametric 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 first step achieves balance, even though subsequent parametric models are not well specified; or if balancing is incorrect, while parametric models are well specified (Ho et al., 2007).

4.1 Preprocessing data: entropy balancing

Entropy balancing 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 the treatment effect. For example, we expect the comprehensive matches to grammar pupils to display higher average cognitive ability scores at age 11 than secondary modern pupils and their respective comprehensive matches. Upon surveying a range of matching procedures, entropy balancing was found to achieve the best balance among the covariates of interest, while retaining all important information from

[8]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.

[9]Validity of matching estimators relies on the conditional independence assumption, expressed as $Y_i^j \perp S_i|\mathbf{X}_i$, with $j = 0, 1$. In the present case however, we do not explicitely make this assumption, since we are not interested in treatment effect directly estimated via matching, but rather in the weights for untreated individuals resulting from this procedure.

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

the original sample[11]. Developed by Hainmueller (2012), the procedure 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[12].

The candidate covariates for the balancing procedure are selected on the basis of their expected relationship to both treatment and outcomes (Caliendo and Kopeinig, 2008)[13]. 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 as 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[14], so only age 7 cognitive ability scores are selected for balancing (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[15]. Two more background variables are included: mother's interest in child education (classified in four categories) and a dummy for high or middle-high father's SES, both measured when the child is aged 11. By balancing mean, variance and skewness of the five included covariates, as well as their pairwise interactions, we achieved very close balance, without compromising the feasibility of the minimization procedure required for the entropy balancing. Figure 1 shows density kernel estimates for the three ability measures before and after balancing, separately for the GC and SMC samples. In both samples, applying balancing weights to comprehensive pupils yields a density that resembles more closely that of the treated, thus strengthening confidence in our ability of estimating an unbiased treatment effect.

A final note on balancing is in order. The concern around comparability issues between selective and non-selective areas could presumably be addressed by including local area characteristics in the balancing procedure. However, when looking at the density kernel estimates for these, we find that for most characteristics there is a very large overlap

[11]Alternatives surveyed included propensity score matching, and a combination of coarsened exact matching followed by propensity score matching (Iacus et al., 2012; 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.

[12]All empirical analysis is conducted using Stata15. The Stata package ebalance allows for a straightforward implementation of the entropy balancing algorithm (Hainmueller and Xu, 2013).

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

[14]Coaching effects reflect 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).

[15]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 carry over to this variable.

between the characteristics of treated and control pupils in both the GC and SMC sample, which decreases concerns around area-level differences biasing our treatment effect (see Figures 2 and 3). As a sensitivity check, we include fourteen area characteristics in the entropy balancing algorithm. Balance achieved is reasonably good for all individual and regional characteristics in both samples, but our main findings are not affected, and we therefore proceed with the simpler balancing algorithm in our main specification.

4.2 Parametric regressions

We use the weights obtained via entropy balancing for the control units in parametric regressions where we control for a larger set of pre-treatment covariates that might affect our outcomes of interest, in addition to the five key covariates used in the balancing procedure. Assuming for each sample $j = GC, SMC$ a constant treatment effect α^j for all individuals, denoting two different treatment effects in the two estimation samples, we estimate the following by ordinary least squares (OLS):

$$Y_i^j = \beta_0 + \alpha^j S_i^j + \beta_1 C_i + \beta_2 NC_i + \beta_3 B_i + \beta_4 SES_i + \beta_5 HE_i + \beta_6 PS_i + \beta_7 LA_i + \epsilon_i^j, \quad (6)$$

with constant β_0 and the binary treatment variable S_i equal to 1 for grammar attendance in the GC sample, or for secondary modern attendance in the SMC sample, and 0 for comprehensive attendance. Covariates are cognitive and non-cognitive skills, C_i and NC_i, the vector of individual background characteristics B_i, including sex and ethnicity, family socioeconomic status SES_i, childhood health endowment HE_i (to rule out reverse causality of health on schooling), primary school and local authority characteristics PS_i and LA_i, while ϵ_i is a random error term.

In a second specification, we include interactions 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. We estimate

$$Y_i^j = \gamma_0 + \alpha_1^j S_i^j + \sum_{q=2}^{4} \alpha_{2q}^j S_i^j \times C_{iq} + \sum_{q=2}^{4} \alpha_{3j}^j S_q^j \times NC_{iq} + X'\gamma + \upsilon_i^j \quad (7)$$

by OLS, where for ease of notation $X'\gamma$ is the vector of all individual characteristics as in Equation (6) and respective coefficients. The estimated coefficients α_{2q}^j and α_{3q}^j will then reflect the effect of grammar attendance, say, compared to comprehensive, for pupils in each quartile q of the ability distribution, with the interaction with the lowest quantile as the base category.

Given the NCDS cohort entered secondary school in 1969, only four years after the Labour-backed Circular 10/65, we estimate a third specification distinguishing between comprehensive schools that were formerly grammar or secondary modern, versus comprehensive that are either purpose-built or amalgamated. This is to ensure that the effect estimate of school type is not confounded by comprehensive schools still transitioning from their grammar or secondary modern origin. Within the GC sample, we set grammar as the base category and two treatment variables: one being an indicator for attendance to a comprehensive that is a former grammar CF_i, and one indicating attendance to a newly built comprehensive CO_i. A similar approach is then implemented to the SMC sample too. We estimate

$$Y_i^j = \eta_0 + \delta_1^j CF_i + \delta_2^j CO_i + X'\eta + \xi_i^j \tag{8}$$

by OLS, where $X'\eta$ is again the vector of individual characteristics, as in Equation (6), and respective coefficients. We now distinguish between three types of school, which allows us to distinguish between the effect of attending a comprehensive that used to be a grammar (or secondary modern) ($\hat{\delta_1^j}$) and a purpose-built one ($\hat{\delta_2^j}$), compared to attending grammar (or secondary modern).

4.3 Robustness checks

In the choice of the empirical strategy for estimation of Equation (3), we consider the possibility that, entropy balancing notwithstanding, treatment assignment S_i could be endogenous due to omitted variables or unobservables, a classic problem in the literature on returns to education[16]. Since OLS with endogenous treatment yields biased and inconsistent estimates, we conduct some checks in order to find the estimation method most likely to avoid bias. The standard tool to solve endogeneity is to implement an instrumental variable (IV) strategy. As mentioned, the literature has used instruments such as share of comprehensive schools in the LEA at time of schooling (Basu et al., 2018), and political control in the area (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 school type[17].

The instrument Z_i satisfies the relevance requirement $corr(Z_i, S_i) \neq 0$, as LEA % of comprehensive pupils proxies supply of comprehensive places in the LEA, so both grammar and secondary modern attendance are expected to be significantly and negatively correlated with Z_i. Secondly, the validity of the exclusion restriction assumption requires

[16]For any given outcome of interest

$$Y_i^j = \mathbf{X}_i\beta + \epsilon_i, \tag{9}$$

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

[17]Percentage of grammar and of secondary modern as shares of total LEA schools were available, but not as precise as share of pupils when it comes to proxying supply of places. Percentage of grammar and secondary modern pupils in each LEA was not available from the sources mentioned.

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 Equation (6). Average LEA population in 1971 was 413,649 (Registrar General for England and Wales, 1971), and this large size makes it unlikely that LEA-level characteristics could determine individual's behaviours and outcomes, as peer effects and environmental factors are weaker in larger areas than say, a neighbourhood or a school[18]. The models for the outcomes of interest are estimated by Two Stage Least Squares (2SLS)[19]. 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. The second stage, counterpart to Equation (3), uses the school type predicted in the first as a regressor for the outcome equation.

We base our choice of strategy on the the following criteria. First, a rule of thumb is that when confidence intervals for IV estimators contain OLS estimates, it is always advisable to use OLS. We find this to be the case for all of our outcomes. Second, under treatment exogeneity, OLS has superior finite sample properties to IV estimators, as well as smaller variance (Sargan, 1958). We therefore conduct Durbin-Wu-Hausman tests of endogeneity of school type for all outcomes of interest, before and after balancing, including all available controls. For all our outcomes, the test fails to reject the null of exogeneity (results available in Appendix table A1). On the basis of this evidence, we keep 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. A recent paper reviewing several empirical IV applications also argued that OLS performs better in the presence of a non-random error term, and that evidence of OLS being substantially biased in applications is scarce at best (Young, 2018). Nevertheless, as a robustness check, we implement IV strategies after preprocessing the sample via entropy balancing, and leave results in section 5.3 for the interested reader.

A further robustness check of our approach 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 both post-secondary school maths test scores and pre-secondary school scores. In theory, we would not expect secondary school type to be a significant predictor of scores prior to treatment, unless the model is misspecified or the estimation strategy unable to prevent bias. In some recent work, Basu et al. (2018) conduct a similar

[18]We further control for several LEA-level characteristics, such as county proportion of unemployed, council tenants, house owners and professional categories for household heads, in the parametric specification of all outcome regressions.

[19]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 not included in the main paper. See Terza et al., 2008 for more details.

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 at ages 11 and 16, as a measure of general health[20]. If the balanced 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 we are successfully controlling for pre-existing differences in the compared samples. Results are discussed in section 5.4.

5 Results

5.1 Characteristics by type of school

All individual characteristics of interest included in the models are measured prior to starting secondary school (see Table A2 in the Appendix describing average characteristics by type of school). Future grammar, comprehensive and secondary modern pupils differ most notably in the three measures of ability. At both age 7 and 11, grammar pupils present highest cognitive ability, the main determinant of entry test success, followed by comprehensive and then by secondary modern pupils. Grammar pupils also display higher non-cognitive abilities, as well as higher proportions of female and first-born, and are less likely to have two or more siblings. On average grammar pupils are more advantaged, both in terms of parental interest in their education and socioeconomic background, and they are much more likely to state at 11 that they plan to study after compulsory schooling. Grammar pupils have slightly higher health endowment than the other two categories of pupils, with lower probabilities of mothers smoking during pregnancy and 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.

Table 2 summarises the outcomes. On average grammar pupils display higher well-being, while secondary modern students fare worst out of the three groups, 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. Average hourly wages are significantly higher for grammar pupils, while lower for comprehensive and lowest for secondary mod-

[20]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.

ern pupils. Age 50 average hourly wages are almost twice as high as age 33 ones, reflecting higher expertise but also inflation over time. The probability of being employed at both ages is slightly higher for grammar pupils than the rest, who display similar proportions of employed. Grammar pupils also report better health, while secondary modern students do worst out of the three groups. A similar pattern is observed with biometric measures, all 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.

5.2 Main results

We summarise the key contribution of entropy balancing in Tables 3 and 4. These show the first three moments of the five key covariates of interest, before and after balancing, for the GC and SMC samples respectively. The upper panel in each table shows raw mean, variance and skewness, while the lower panel shows re-weighted moments for control individuals from the comprehensive sample, using 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 mean, variance and skewness, increasing confidence that the joint distribution of these variables will be more similar in the two groups after matching.

Tables 5 to 10 report results for the main outcome regressions of interest, all estimated by OLS for the matched GC and SMC samples separately. All continuous variables are standardised for ease of comparison except for logged hourly wage, which can be interpreted in terms of percentages. All models for binary variables are estimated via probit regressions, and we show marginal effects in the tables. Although we only show the coefficients on the treatment, sex and five key variables, all models are estimated controlling for all covariates described in Table A2 (full results available upon request).

Table 5 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 0.2 SD in the school aspirations of grammar pupils), but with lower life satisfaction at age 33 (0.13 SD), compared to comprehensive attendance. The ability variables, on the other hand, all display a significant association with selected outcomes. A 0.10 increase in cognitive skills at age 7 is linked to a 0.06 SD increase in self-efficacy at 33. A 0.10 increase in non-cognitive skills is associated with higher positive feelings about school at 16 (0.08 SD) and higher life satisfaction and self-efficacy at 33 (0.12 SD and 0.13 SD respectively), as well as lower probability of drug use (0.03 SD). Higher relative cognitive ability is significantly linked to higher positivity about school at age 16, as well as higher job positivity at 33. Magnitudes of these effects are 0.15 and 0.08 SD respectively for a 10 percentage point increase in relative cognitive ability (sufficient to be shifted to the next upper decile of the cognitive ability distribution). The likelihood of displaying positive work aspirations at 16 for a 100% increase in relative cognitive ability increases up to 18%. Interestingly, relative cognitive ability is also significant at 10% as a predictor of higher drug use.

Table 6 shows estimation results for labour outcomes for the matched GC sample. Grammar is only significant at 10% for age 50 log-transformed wage, and on average it raises hourly wage by 9% compared to attending comprehensive. Relative cognitive ability is a significant predictor of wage at both ages considered, and coefficients are large: a 0.10 increase in the cognitive ability ranking raises hourly wages by approximately 6.5% at age 33 and 8% at age 50. Non-cognitive ability on the other hand, only appears significant for age 50 wages (8% increase for a 0.10 increase in childhood non-cognitive skills). Being female significantly reduces expected wage and the probability of being in employment at both ages.

Table 7 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% (magnitude 0.1 SD). Again cognitive abilities present striking results. Cognitive ability at age 7 is negatively and significantly related to age 45 BMI levels, as well as to cholesterol ratio and tryglicerides, two biomarkers for risk of CVD (magnitudes reaching 0.06 SD for a 0.10 increase in the cognitive index). 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 is associated with 0.05 and 0.04 SD decreases in the risk biomarkers respectively. 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 0.24 SD (above 1 point) lower and they score lower in all biomarkers for risk, except for CRP.

Table 8 shows results for well-being outcomes, estimated using the matched SMC sample. Secondary modern attendance increases positive feelings about school at 16 (0.1 SD) compared to comprehensive. Pre-secondary schooling non-cognitive skills are once more significantly correlated with well-being, and they are positively and significantly related to all of our positive outcomes. Associations range between 0.04 SD (for self-efficacy) and 0.06 SD (for school aspirations and life-satisfaction) for a 0.10 increase in non-cognitive skills. For the binary variables, a 100% increase in childhood non-cognitive skills is linked to 14% higher probabilities of positive work aspirations, 13% lower probability of crime and 24% lower probability of drug use. 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 0.04 and 0.13 SD, depending on the outcome.

Table 9 displays results for labour outcomes. All else equal, secondary modern attendance increases average wages at ages 33 and 50 by roughly 5% and 8.5% respectively, and the probability of being employed at 33 by 3%. Non-cognitive skills are significantly and positively related to the probability of being employed at ages 33 and 50, although effects are small, and to age 33 wages (2.5% for every 0.10 increase in non-cognitive skills). A 0.10 increase in cognitive ability ranking increases hourly wages significantly by approximately 3.7% at age 33 and 4.4% at age 50, and the probability of employment at age 50 (1%).

Finally, Table 10 shows results for health outcomes, estimated for the matched SMC 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 an approximate 3.5% increase in the probabilities of scoring high self-assessed health and low malaise at age 50, as well as a reduction of 0.06 SD in mental health problems at 45. Although coefficients are small, this is a strong result, and compared to the GC sample it indicates that higher non-cognitive skills might have a protective role for health for pupils of lower average cognitive ability, such as it is the case in the SMC sample. Relative position by cognitive ability is also related to increases in the probability of high SAH and low malaise (magnitudes of effects of 0.02 and 0.01 SD 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.

Estimation of the specifications illustrated by Equations (7) and (8) did not add any further insights to our main message. Interacting treatment with high and low levels of cognitive and non-cognitive skills, as well as considering the effect of comprehensive schools that were formerly grammar or secondary modern separately from other types of comprehensive schools did not produce significantly different results from our main specification (we leave results in Appendix Tables A5-A16).

5.3 IV results

Tables A17 to A21 in the Appendix display results for 2SLS estimation of the IV models used as a robustness check for our main results. Table A17 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 (as set out in section 4.3). 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[21] (Stock et al., 2002).

The four 2SLS results tables show that significant effects of grammar and secondary modern found with OLS estimates are not always matched by IV estimates. This might be due to standard errors associated to IV estimates of treatment coefficients being at least twice as large as OLS standard errors. However, as noted above, OLS estimates of treatment effect fall within the confidence intervals for IV estimates for all outcomes, meaning there is no statistically significant difference between the two. Note also that 2SLS coefficients for the ability variables (not instrumented) are largely similar to OLS ones in magnitude and significance in all four of our IV results tables[22].

5.4 Falsification test results

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 our empirical strategy. There are some key differences to 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 balanced samples followed by parametric regressions 'pass' the falsification test, meaning that the coefficient on comprehensive is not significant for outcomes prior to secondary school, then this would provide support for 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. Results for age 11 and age 16 scores confirm what was found by the original authors (see Appendix Tables A24 and A26). Comprehensive attendance, used as treatment for both groups for comparability with the original test, is a significant and negative predictor of both age 11 and age 16 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,

[21] 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 (Angrist and Pischke, 2009).

[22] 2SRI results, here not shown, are similar to OLS results, although the magnitude of coefficients varies in several cases. Generalised residuals saved from the first stage of 2SRI 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. Again, we refer to Terza et al., 2008 for more details.

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 (see Appendix Tables A25 and A27). 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 our empirical strategy for health outcomes, while they suggest some caution 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 entropy balancing, an intuitive and effective matching method, with parametric regression, which yields doubly robust estimates, thus helping create a quasi-experimental setting to evaluate an educational reform with no clear rollout. Criticisms previously advanced by Manning and Pischke (2006) in this literature targeted value-added methodologies and IV regressions used to explore the effects of selective schooling on educational achievement. These were 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. 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, with the exception of some labour outcomes. In their analysis of Scottish data, Del Bono and Clark (2016) find no significant effects for most of the adult outcomes considered, except for female income and fertility. Dustmann et al. (2016) find no significant differences by middle school track for long-term education and labour outcomes in Germany. With respect to health effects in the English context, Basu et al. (2018) find no significant effects for self-assessed health and smoking, and we extend their result by adding evidence on biomarkers for risk of CVD and well-being outcomes. 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. In our application however, we do not find that accounting for heterogeneity changes our main findings. Interestingly, we find that in the short-term pupils who attended grammar and secondary modern display a more positive attitude towards school than comprehensive pupils. This effect may translate into higher wages and better employment prospects later over the life time. 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 and well-being outcomes. We also find that childhood cognitive abilities can be important determinants of health, well-being and labour outcomes later in life, even when accounting for non-cognitive skills, which agrees with the literature on the economics of human capital (Auld and Sidhu, 2005, Conti and Heckman, 2010, Bijwaard et al., 2015). An interesting feature of our findings is that they also suggest that the importance of non-cognitive skills for life outcomes may vary depending on the level of cognitive skills. This is the case for the protective role of non-cognitive skills for health, which emerges as significant only in the lower cognitive ability sample.

A reason why we do not find a consistently significant treatment effect on health and well-being could be that selection on ability in secondary schooling does not affect the channels that are assumed to lead to better adult 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 resources towards early childhood education interventions, rather than new selective schools. On the other hand, channels that affect these outcomes might also be formed later on, after secondary schooling. This is the case of changes produced in adulthood via University attendance, career path, work and residential environment and so on. Research on the mediators between education and health and well-being, as well as on the key life stages at which these mediators are affected, could inform us more on the potential beneficial effects of educational policy for positive adult outcomes (Kautz et al., 2014).

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 comprehensive, mixed-ability system, allowing us to explore health and well-being effects at several points of these individuals' lives over time. The data is first preprocessed through entropy balancing, 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. The only exception are some labour outcomes, which are better for selective pupils, and short-term schooling aspirations, which may be linked. Childhood cognitive and non-cognitive ability measured prior to secondary schooling, on the other hand, play a significant role as predictors of later health and well-being. Their role as either direct causal predictor of well-being or mediators between education and well-being should be the subject of further research to explore the determinants of differences among individual outcomes, which would be informative for future educational policy.

References

Angrist, J. D. (1998). 'Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants'. *Econometrica* vol. 66 (2), pp. 249–288.

Angrist, J. D. and Pischke, J. S. (2009). *Mostly Harmless Econometrics*. Princeton University Press.

Auld, M. C. and Sidhu, N. (2005). 'Schooling, cognitive ability and health'. *Health Economics* vol. 14 (10), pp. 1019–1034.

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

Benzeval, M. et al. (2014). 'Biomarker User Guide and Glossary'. Understanding Society: The UK Household Longitudinal Study.

Bijwaard, G. E., Kippersluis, H. V., and Veenman, J. (2015). 'Education and health : The role of cognitive ability'. *Journal of Health Economics* vol. 42, pp. 29–43.

Bolton, P. (2017). 'Grammar School Statistics'. House of Commons Briefing Paper No. 1398.

Bonhomme, S. and Sauder, U. (2011). 'Recovering distributions in difference-in-differences models : a comparison of selective and comprehensive schooling'. *Review of Economic Studies* vol. 93 (2), pp. 479–494.

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

Brunello, G. and Checchi, D. (2007). 'Does School Tracking Affect Equality of Opportunity? New International Evidence'. *Economic Policy* vol. 22 (52), pp. 783–861.

Burgess, S. (2016). 'Human Capital and Education: The State of the Art in the Economics of Education'. IZA Discussion Paper No. 9885.

Burgess, S., Crawford, C., and Macmillan, L. (2017). 'Assessing the role of grammar schools in promoting social mobility'. DQSS Working Paper No. 17/09, UCL Institute of Education.

Burgess, S., Dickson, M., and Macmillan, L. (2014). 'Selective Schooling Systems Increase Inequality'. DQSS Working Paper No. 14/09, UCL Institute of Education.

Caliendo, M. and Kopeinig, S. (2008). 'Some practical guidance for the implementation of propensity score matching'. *Journal of Economic Surveys* vol. 22 (1), pp. 31–72.

Campbell, F. et al. (2014). 'Early Childhood Investments Substantially Boost Adult Health'. *Science* vol. 343 (6178), pp. 1478–1485.

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

Cawley, J. et al. (1996). 'Measuring the Effects of Cognitive Ability'. NBER Working Paper No. 5645.

Comprehensive School Committee (1971). 'Comprehensive Education Statistics'. 1971 Bulletin of the Comprehensive Schools Committee.

Conti, G. and Heckman, J. J. (2010). 'Understanding the Early Origins of the Education–Gradient: A Framework That Can Also Be Applied to Analyze Gene–Environment Interactions'. *Perspectives on Psychological Science* vol. 5 (5), pp. 585–605.

Cutler, D. M. and Lleras-Muney, A. (2011). 'Understanding Differences in Health Behaviors by Education'. *Journal of Health Economics* vol. 29 (1), pp. 1–28.

Dearden, L., Ferri, J., and Meghir, C. (2002). 'The Effect of School Quality on Educational Attainment and Wages'. vol. 84 (1), pp. 1–20.

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

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

Fuchs, V. R. (1982). 'Time Preference and Health: An Exploratory Study'. *Economic Aspects of Health*. Ed. by V. R. Fuchs. University of Chicago Press, pp. 93–120.

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

Grossman, M. (1972). 'On the Concept of Health Capital and the Demand for Health'. *Journal of Political Economy* vol. 80 (2), pp. 223–255.

Hainmueller, J. (2012). 'Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies'. *Political Analysis* vol. 20 (1), pp. 25–46.

Hainmueller, J. and Xu, Y. (2013). 'Ebalance: a Stata package for entropy balancing'. *Journal of Statistical Software* vol. 54 (7), pp. 1–18.

Hanushek, E. A. and Wößmann, L. (2006). 'Does Early Tracking Affect Educational Inequality and Performance? Differences-in-Differences Evidence across Countries'. *Economic Journal* vol. 116, pp. C63–C76.

Heckman, J. J., Smith, J., and Clements, N. (1997). 'Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts'. *The Review of Economic Studies* vol. 64 (4), pp. 487–535.

Ho, D. E. et al. (2007). 'Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference'. *Political Analysis* vol. 15 (3), pp. 199–236.

Iacus, S. M., King, G., and Porro, G. (2012). 'Causal Inference without Balance Checking : Coarsened Exact Matching'. *Political Analysis* vol. 20 (1), pp. 1–24.

Imbens, G. W. (2004). 'Nonparametric estimation of average treatment effects under exogeneity: a review'. *Review of Economics & Statistics* vol. 86 (1), pp. 4–29.

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

Jones, A. M., Rice, N., and Rosa Dias, P. (2012). 'Quality of schooling and inequality of opportunity in health'. *Empirical Economics* vol. 42 (2), pp. 369–394.

Kautz, T. et al. (2014). 'Fostering and Measuring Skills : Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success'. OECD Education Working Paper No. 110.

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

— (1996). *Going Comprehensive in England and Wales: a study of uneven change.* London: Routledge.

Leuven, E. and Sianesi, B. (2012). 'PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing'. Statistical Software Components S432001, Boston College Department of Economics.

Long, R., Foster, D., and Roberts, N. (2017). 'Grammar schools in England'. House of Commons Briefing Paper No. 7070.

Mackinnon, J. G. and Davidson, R. (2003). *Econometric Theory and Methods.* Oxford University Press.

Manning, A. and Pischke, J. S. (2006). 'Comprehensive versus selective schooling in England in Wales: What do we know?' NBER Working Paper No. 12176.

Power, C. and Elliott, J. (2006). 'Cohort profile: 1958 British birth cohort (National Child Development Study)'. *International Journal of Epidemiology* vol. 35 (1), pp. 34–41.

Registrar General for England and Wales (1971). '1971 Census aggregate and Local Authority data'. UK Data Service.

Richardson, H. (2016). *Grammar schools: What are they and why are they controversial?* BBC News 8 September. URL: `http://www.bbc.co.uk/news/education-34538222`.

Rubin, D. (1974). 'Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies'. *Journal of Educational Psychology* vol. 66 (5), pp. 688–701.

Rubin, D. B. and Thomas, N. (1996). 'Matching Using Estimated Propensity Scores: Relating Theory to Practice'. *Biometrics* vol. 52 (1), pp. 249–264.

Sargan, J. D. (1958). 'The Estimation of Economic Relationships using Instrumental Variables'. *Econometrica* vol. 26 (3), pp. 393–415.

Stock, J. H., Wright, J. H., and Yogo, M. (2002). 'A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments'. *Journal of Business & Economic Statistics* vol. 20 (4), pp. 518–529.

Terza, J. V., Basu, A., and Rathouz, P. J. (2008). 'Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling'. *Journal of Health Economics* vol. 27 (3), pp. 531–543.

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

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

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

— (2012). 'National Child Development Study: Sweep 8, 2008-2009.' [data collection]. 3rd Edition. UK Data Service. SN: 6137.

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

World Health Organisation (2017). *Body mass index - BMI.* URL: `http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi` (visited on 09/24/2017).

Young, A. (2018). 'Consistency without Inference : Instrumental Variables in Practical Application'. mimeo, London School of Economics.

Figures and tables

Figure 1: Density kernel estimates for ability measures, for the GC sample (top row) and the SMC sample (bottom row). The dashed line illustrates desnity kernels for comprehensive pupils, balanced with the weights obtained via entropy matching so that they are more comparable to treated individuals.

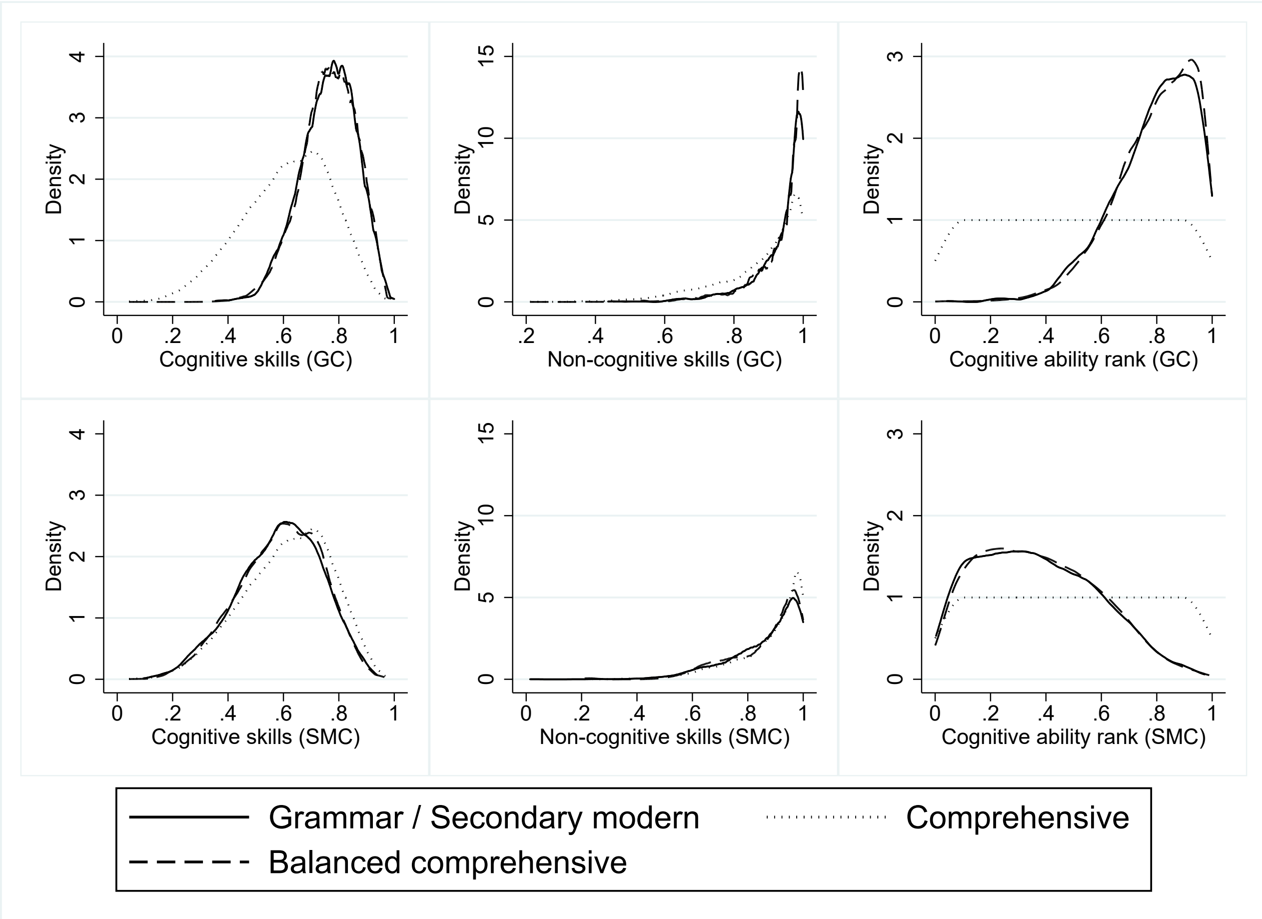

Figure 2: Density kernel estimates for local area characteristics from 1971 census, for grammar, comprehensive and comprehensive reweighted via entropy balancing weights.

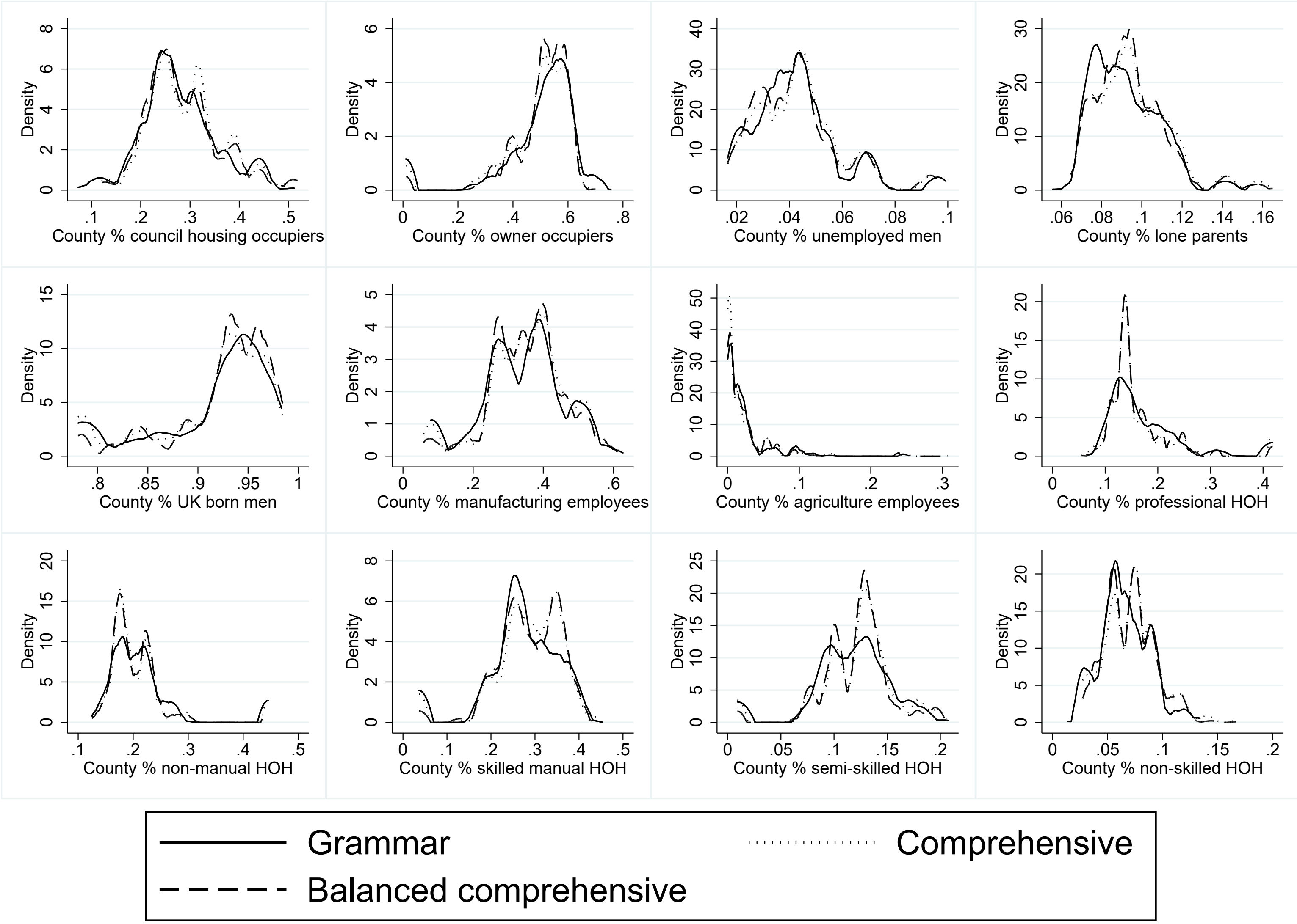

Figure 3: Density kernel estimates for local area characteristics from 1971 census, for secondary modern, comprehensive and comprehensive reweighed via entropy balancing weights.

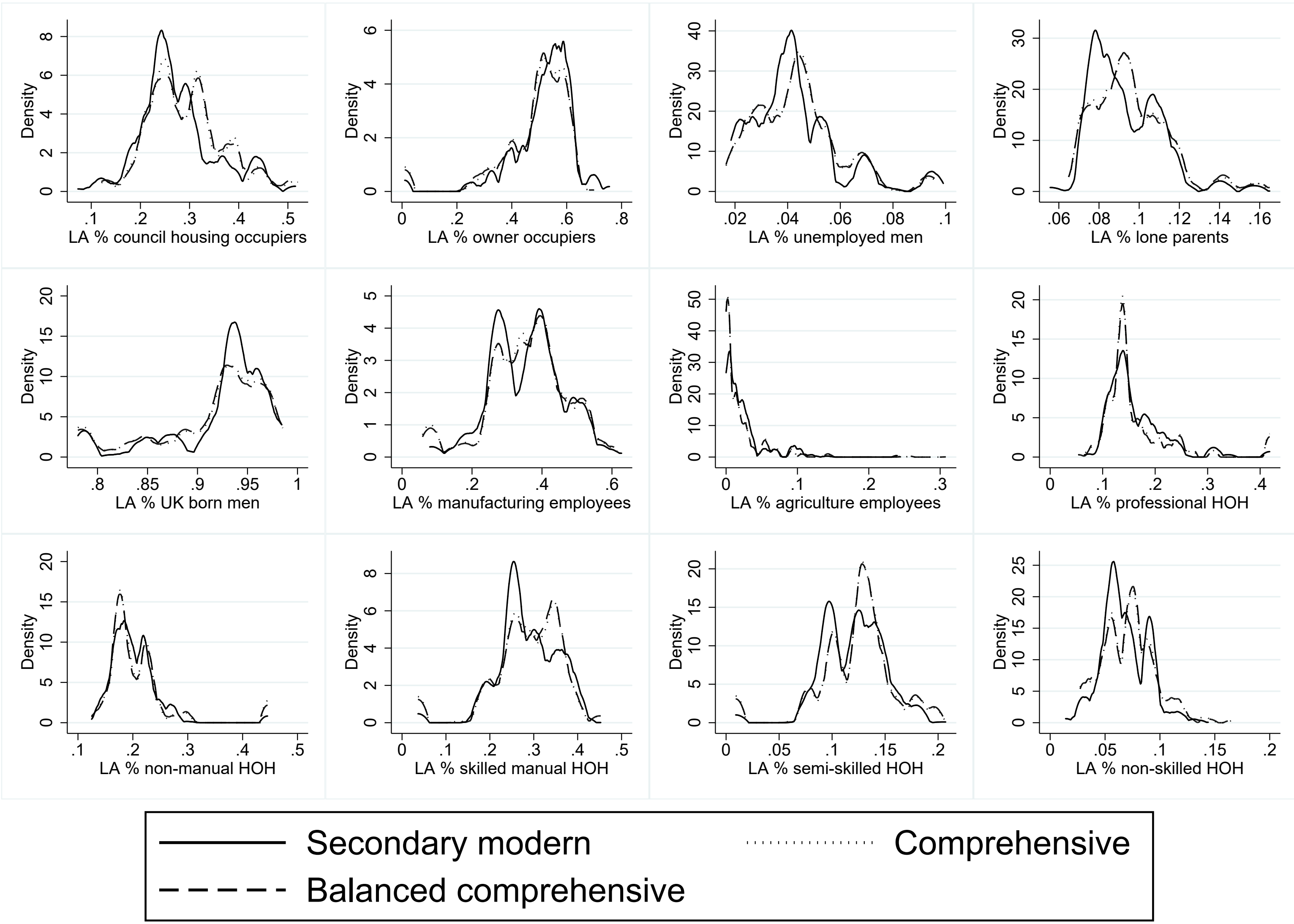

Table 1: Summary statistics: school characteristics by type of secondary school attended. Source: NCDS wave 3.

	Grammar				**Comprehensive**				**Secondary modern**			
	Mean	s.d.	Min	Max	Mean	s.d.	Min	Max	Mean	s.d.	Min	Max
Number of pupils	658.57	191.52	68	1680	1107.76	377.94	174	2674	670.33	283.41	80	1750
Pupil teacher ratio	16.05	1.55	2	20	17.11	1.86	7	45	18.17	1.90	7	41
No. pupils passing 2+ A-levels	45.28	23.24	0	140	16.63	22.26	0	155	0.70	2.86	0	37
No. pupils doing degree	26.31	16.33	0	89	9.54	14.26	0	112	0.31	1.66	0	26
Single-sex school	0.68	0.47	0	1	0.13	0.34	0	1	0.26	0.44	0	1
Lacks library	0.25	0.43	0	1	0.21	0.41	0	1	0.23	0.42	0	1
Lacks science labs	0.23	0.42	0	1	0.19	0.39	0	1	0.35	0.48	0	1
Lacks sport facilities	0.32	0.47	0	1	0.34	0.47	0	1	0.38	0.49	0	1
English class size	26.99	4.76	4	40	25.88	5.48	1	44	25.45	5.62	1	50
Maths class size	25.61	5.09	2	40	25.68	5.78	1	46	25.13	5.93	5	42
Regular physical punishment	0.24	0.43	0	1	0.41	0.49	0	1	0.42	0.49	0	1
Has a parent-teacher association	0.78	0.42	0	1	0.74	0.44	0	1	0.52	0.50	0	1
Observations	1314				6135				2710			

Table 2: Descriptive statistics of wellbeing and health outcomes by type of secondary school attended. School aspirations is a variable constructed via PCA, grouping cohort members' answers to five related questions, measuring the individual's attitude towards school and studying. Work aspirations 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. For the wage outcome, we excluded from the analysis 13 individuals with weekly income above £10,000. The self-assessed health scale goes from 1)excellent health to 5)very poor health.

	Grammar				**Comprehensive**				**Secondary modern**			
	Mean	s.d.	Min	Max	Mean	s.d.	Min	Max	Mean	s.d.	Min	Max
Well-being measures												
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
Labour outcomes												
Hourly wage at 33	9.33	12.39	0	148	7.19	12.52	0	357	6.42	10.86	0	300
Employed at 33	0.84	0.37	0	1	0.79	0.40	0	1	0.80	0.40	0	1
Hourly wage at 50	22.30	30.09	0	462	16.33	12.91	0	235	15.16	10.97	0	85
Employed at 50	0.92	0.27	0	1	0.86	0.35	0	1	0.85	0.36	0	1
Survey health measures												
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
Biometric health measures												
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
Observations	1308				6002				2651			

Table 3: Pre- and post-matching moments of key covariates for the GC sample. Mean, variance and skewness of the pairwise interactions of the five covariates listed are also balanced (not shown).

	Grammar N=1040			**Comprehensive N=4663**		
	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. Mean, variance and skewness of the pairwise interactions of the five covariates listed are also balanced (not shown).

	Secondary modern N=1991			**Comprehensive N=4663**		
	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 wellbeing outcomes (balanced grammar and comprehensive sample). For probit models, marginal effects are displayed.

	School	**Work**	**Life sat.**	**Self-eff.**	**Job positiv.**	**Crime**	**Drugs**
Grammar	0.1977***	0.0130	-0.1307*	-0.0535	-0.0014	-0.0142	-0.0103
	(0.0425)	(0.0141)	(0.0559)	(0.0491)	(0.0475)	(0.0202)	(0.0213)
Cognitive skills	0.1648	-0.0625	-0.0561	0.5919*	0.2390	0.0706	-0.0051
	(0.2212)	(0.0733)	(0.2946)	(0.2587)	(0.2500)	(0.1055)	(0.1108)
Non-cognitive skills	0.7552**	0.0615	1.1847**	1.1133***	0.0421	-0.0450	-0.3070*
	(0.2778)	(0.0899)	(0.3798)	(0.3331)	(0.3211)	(0.1266)	(0.1296)
Relative cogn. ability	1.4445***	0.1832***	-0.1145	0.3567+	0.8344***	-0.0783	0.1491+
	(0.1571)	(0.0497)	(0.2097)	(0.1841)	(0.1782)	(0.0750)	(0.0813)
Female	0.0046	0.0101	0.0641	-0.0683	-0.3803***	-0.1101***	-0.0624**
	(0.0425)	(0.0141)	(0.0563)	(0.0493)	(0.0477)	(0.0201)	(0.0211)
Mother's interest	0.1051***	0.0020	0.0566	-0.0112	0.0640+	0.0230	0.0205
	(0.0284)	(0.0091)	(0.0383)	(0.0336)	(0.0329)	(0.0146)	(0.0151)
Father's SES	0.0282	0.0188	0.0823	-0.0363	0.0601	-0.0791**	-0.0055
	(0.0631)	(0.0196)	(0.0835)	(0.0733)	(0.0712)	(0.0280)	(0.0315)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4044	4156	3131	3083	3145	3277	3279
F	8.1650		1.2206	1.7262	4.1799		
chi2		52.28				63.44	65.97

Standard errors in parentheses. Father's SES is a dummy for high/middle-high SES. All controls included.

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

Table 6: Models for labour outcomes (balanced grammar and comprehensive sample). For probit models, marginal effects are displayed.

	Log hourly wage 33	**Employed at 33**	**Log hourly wage 50**	**Employed at 50**
Grammar	0.0587	0.0319	0.0881+	0.0199
	(0.0400)	(0.0197)	(0.0528)	(0.0171)
Cognitive skills	0.0100	0.0798	-0.0284	0.0711
	(0.2072)	(0.1039)	(0.2739)	(0.0877)
Non-cognitive skills	0.4189	0.0813	0.7619*	0.0928
	(0.2691)	(0.1387)	(0.3647)	(0.1131)
Relative cogn. ability	0.6553***	0.0601	0.7929***	-0.0116
	(0.1518)	(0.0725)	(0.2049)	(0.0636)
Female	-0.4287***	-0.2651***	-0.2218***	-0.0652***
	(0.0400)	(0.0231)	(0.0542)	(0.0183)
Mother's interest	-0.0086	-0.0004	0.0558	0.0056
	(0.0275)	(0.0138)	(0.0374)	(0.0118)
Father's SES	0.0253	0.0390	-0.0520	0.0182
	(0.0596)	(0.0283)	(0.0804)	(0.0250)
Controls	Yes	Yes	Yes	Yes
Observations	2460	3323	1551	2852
F	5.2696		2.7162	
chi2		189.20		47.44

Standard errors in parentheses. Father's SES is a dummy for high/middle-high SES. All controls included.

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

Table 7: Models for health outcomes (balanced grammar and comprehensive sample). For probit models, marginal effects are displayed.

	High SAH	**Low malaise**	**Mental ill-health**	**BMI**	**Chol ratio**	**Triglycerides**	**CRP**	**Fibrinogen**
Grammar	-0.0055	0.0170	0.0122	-0.1056+	0.0385	-0.0069	0.0031	0.0133
	(0.0287)	(0.0231)	(0.0524)	(0.0582)	(0.0581)	(0.0564)	(0.0532)	(0.0612)
Cognitive skills	0.2267	0.0699	-0.1821	-0.6314*	-0.5410+	-0.5765*	-0.1312	-0.0597
	(0.1496)	(0.1201)	(0.2706)	(0.2988)	(0.2981)	(0.2887)	(0.2712)	(0.3121)
Non-cognitive skills	0.2213	0.0946	-0.6169+	-0.4083	-0.3525	-0.2488	-0.2157	-0.5401
	(0.1908)	(0.1557)	(0.3478)	(0.3846)	(0.3853)	(0.3743)	(0.3529)	(0.4062)
Relative cogn. ability	0.0948	0.1047	0.2404	0.0119	-0.2163	-0.0790	-0.4670*	-0.4220+
	(0.1079)	(0.0850)	(0.1978)	(0.2189)	(0.2210)	(0.2141)	(0.2023)	(0.2336)
Female	0.0193	-0.0961***	0.2175***	-0.2428***	-0.8514***	-0.6756***	0.0074	0.2345***
	(0.0289)	(0.0232)	(0.0526)	(0.0581)	(0.0585)	(0.0569)	(0.0536)	(0.0617)
Mother's interest	0.0061	0.0075	0.0151	-0.0689+	-0.0270	-0.0558	-0.0202	-0.0689
	(0.0199)	(0.0159)	(0.0377)	(0.0418)	(0.0424)	(0.0412)	(0.0389)	(0.0448)
Father's SES	0.0135	-0.0582	-0.0187	-0.1021	-0.0120	-0.1181	-0.0178	-0.0338
	(0.0429)	(0.0363)	(0.0772)	(0.0854)	(0.0881)	(0.0855)	(0.0809)	(0.0931)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2875	2854	2787	2759	2327	2333	2302	2295
F			1.4594	2.0902	7.2168	5.6455	0.5411	1.5607
chi2	52.8400	56.4290						

Standard errors in parentheses. Father's SES is a dummy for high/middle-high SES. All controls included.

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

Table 8: Models for wellbeing outcomes (balanced secondary modern and comprehensive sample). For probit models, marginal effects are displayed.

	School	**Work**	**Life sat.**	**Self-eff.**	**Job positiv.**	**Crime**	**Drugs**
Secondary modern	0.0990**	0.0180	-0.0066	0.0793+	0.0274	-0.0205	-0.0069
	(0.0316)	(0.0155)	(0.0461)	(0.0443)	(0.0438)	(0.0162)	(0.0155)
Cognitive skills	-0.0875	0.0316	0.2017	0.3846*	0.2599	0.0839	0.1074
	(0.1317)	(0.0633)	(0.1968)	(0.1886)	(0.1872)	(0.0689)	(0.0658)
Non-cognitive skills	0.5864***	0.1352*	0.5851**	0.3925*	0.4657*	-0.1251+	-0.2393***
	(0.1342)	(0.0630)	(0.1976)	(0.1899)	(0.1885)	(0.0643)	(0.0609)
Relative cogn. ability	1.2817***	0.4142***	-0.2025	0.2251+	0.6089***	-0.0428	0.0314
	(0.0944)	(0.0477)	(0.1389)	(0.1330)	(0.1320)	(0.0487)	(0.0464)
Female	0.0144	-0.0064	0.1552***	-0.0189	-0.5764***	-0.1984***	-0.0830***
	(0.0311)	(0.0153)	(0.0451)	(0.0433)	(0.0429)	(0.0155)	(0.0151)
Mother's interest	0.1073***	0.0143+	0.0084	0.0223	0.0261	0.0063	0.0097
	(0.0158)	(0.0077)	(0.0231)	(0.0221)	(0.0220)	(0.0081)	(0.0078)
Father's SES	0.1089**	0.0307+	0.0458	0.0821	0.0508	-0.0147	-0.0213
	(0.0371)	(0.0177)	(0.0542)	(0.0520)	(0.0517)	(0.0190)	(0.0180)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4665	4818	3588	3535	3597	3777	3779
F	21.0253		2.1487	3.0606	9.4123		
chi2		275.40				206.80	83.91

Standard errors in parentheses. Father's SES is a dummy for high/middle-high SES. All controls included.
+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table 9: Models for labour outcomes (balanced matched secmodern and comprehensive sample). For probit models, marginal effects are displayed.

	Log hourly wage 33	**Employed at 33**	**Log hourly wage 50**	**Employed at 50**
Secondary modern	0.0489+	0.0344*	0.0847*	-0.0194
	(0.0265)	(0.0165)	(0.0421)	(0.0163)
Cognitive skills	0.1826	0.0408	0.0378	0.1089
	(0.1130)	(0.0704)	(0.1854)	(0.0699)
Non-cognitive skills	0.2492*	0.1400*	-0.0402	0.1827**
	(0.1121)	(0.0706)	(0.1927)	(0.0646)
Relative cogn. ability	0.3748***	0.0565	0.4406***	0.1118*
	(0.0797)	(0.0508)	(0.1289)	(0.0507)
Female	-0.4783***	-0.2472***	-0.2670***	-0.0813***
	(0.0258)	(0.0158)	(0.0415)	(0.0163)
Mother's interest	0.0305*	0.0098	0.0428*	-0.0016
	(0.0133)	(0.0083)	(0.0212)	(0.0080)
Father's SES	0.0931**	0.0543**	0.0234	0.0231
	(0.0306)	(0.0189)	(0.0505)	(0.0188)
Controls	Yes	Yes	Yes	Yes
Observations	2766	3821	1689	3230
F	14.4332		3.5331	
chi2		268.53		110.05

Standard errors in parentheses. Father's SES is a dummy for high/middle-high SES. All controls included.

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

Table 10: Models for health outcomes (matched secondary modern and comprehensive sample). For probit models, marginal effects are displayed.

	High SAH	**Low malaise**	**Mental ill-health**	**BMI**	**Chol ratio**	**Triglycerides**	**CRP**	**Fibrinogen**
Secondary modern	-0.0041	0.0059	-0.0292	0.0375	0.0654	0.0232	-0.0223	-0.0286
	(0.0232)	(0.0199)	(0.0474)	(0.0484)	(0.0497)	(0.0500)	(0.0577)	(0.0518)
Cognitive skills	0.0598	0.0394	-0.1499	-0.0884	-0.0250	-0.0818	-0.2531	-0.0475
	(0.1007)	(0.0858)	(0.2049)	(0.2092)	(0.2156)	(0.2174)	(0.2494)	(0.2236)
Non-cognitive skills	0.3448***	0.3453***	-0.6084**	0.1209	-0.2732	-0.2181	0.1961	0.0890
	(0.0965)	(0.0791)	(0.2022)	(0.2054)	(0.2152)	(0.2163)	(0.2444)	(0.2193)
Relative cogn. ability	0.1510*	0.1221*	-0.1565	-0.1346	-0.1013	-0.0817	-0.1588	-0.4426**
	(0.0701)	(0.0609)	(0.1432)	(0.1461)	(0.1514)	(0.1525)	(0.1752)	(0.1571)
Female	-0.0324	-0.1126***	0.2562***	-0.1673***	-0.6600***	-0.5384***	0.1121*	0.2689***
	(0.0228)	(0.0195)	(0.0466)	(0.0477)	(0.0492)	(0.0495)	(0.0571)	(0.0512)
Mother's interest	0.0266*	-0.0170+	0.0168	-0.0793**	-0.0486+	-0.0068	-0.0114	-0.0373
	(0.0114)	(0.0099)	(0.0238)	(0.0242)	(0.0249)	(0.0251)	(0.0290)	(0.0260)
Father's SES	0.0244	-0.0034	0.0511	-0.1524**	-0.0018	-0.0212	-0.0303	-0.0113
	(0.0277)	(0.0236)	(0.0559)	(0.0571)	(0.0590)	(0.0593)	(0.0683)	(0.0613)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3250	3224	3183	3145	2665	2669	2634	2629
F			3.0529	2.3501	6.3015	4.1947	1.8252	2.4136
chi2	109.2700	96.6975						

Standard errors in parentheses. Father's SES is a dummy for high/middle-high SES. All controls included.

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

A Online Appendix

A.1 Principal component analysis for cognitive ability

Principal component analysis (PCA) finds linear combinations of the variables of interest to explain the maximum variation possible, while reducing data dimensionality. PCA was used to construct a single index of cognitive ability for ages 7, 11 and 16, based on the available tests. Note that in all cases the correlation among the different test scores was high and positive. As a general rule (Kaiser's rule), components are retained if their associated eigenvalue exceeds 1. In all cases, this was only the case for the first component. STATA screeplots of the eigenvalues post-PCA are shown below. Kaiser-Meyer-Olkin measures of sampling adequacy were calculated for the three indices, in order to verify that PCA is indeed appropriate in this case.

Correlation matrix for test scores administered at different ages.

	Maths 7	**Reading 7**	**Copy Design 7**	**Obs.**
Maths 7	1.0000			13,546
Reading 7	0.5425	1.0000		13,576
Copy Design 7	0.3175	0.3377	1.0000	13,525
	Maths 11	**Reading 11**	**General ability 11**	**Obs.**
Maths 11	1.0000			12,810
Reading 11	0.7480	1.0000		12,812
General ability 11	0.8096	0.7457	1.0000	12,813
	Maths 16	**Reading 16**		**Obs.**
Maths 7	1.0000			
Reading 7	0.6552	1.0000		

Cognitive ability indices age 7.

Princ. comp.	**Eigenv.**	**Cum. var. explained**	**Test**	**Fact. loadings**
1	1.81	0.60	Maths	0.61
2	0.73	0.85	Reading	0.62
3	0.46	1.00	Copy designs	0.50

Maths and reading tests have similar factor loadings, while the one associated to copying design test was lower. This 3-part index was preferred anyway, given the values for the Kaiser-Meyer-Olkin sampling adequacy . The first component, with eigenvalue 1.81, explains 0.60 of the variance. Alternatively, the principal component for the two-part index would have eigenvalue 1.54 and explain 0.77 of the variance.

Cognitive ability indices age 11.

Princ. comp.	**Eigenv.**	**Cum. var. explained**	**Test**	**Fact. loadings**
1	2.54	0.85	Maths	0.58
2	0.27	0.94	Reading	0.56
3	0.19	1	General ability	0.58

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. The first principal component, with eigenvalue 2.54, explains 85% of the variance. Note that the three tests have similar loadings associated to them, which supports the idea that the NCDS ability tests can mirror the 11-plus results.

Cognitive ability indices age 16.

Princ. comp.	**Eigenv.**	**Cum. var. explained**	**Test**	**Fact. loadings**
1	1.66	0.83	Maths	0.71
2	0.34	1	Reading	0.71

Both tests have the same factor loading, and the first component, with eigenvalue 1.66, explains 0.83 of the variance.

A.2 Appendix tables

Table A1: Durbin-Wu-Hausman test results for the matched samples. The DWH test allows testing for endogeneity in just-identified models. For each outcome, the null hypothesis is that the treatment variable is exogenous. 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.

GC sample								
	School asp	**Job asp**	**Life sat**	**Self eff**	**Job posit**	**Crime**	**Drugs**	
Hausman test	0.96	0.21	0.07	0.03	0.02	0.17	3.07	
p-value	0.3261	0.6476	0.7856	0.8577	0.8792	0.6837	0.0796	
N	4044	2986	3131	3083	3145	3277	3279	
	Log wage 33	**Employed 33**	**Log wage 50**	**Employed 50**				
Hausman test	1.33	3.6	0.11	0.89				
p-value	0.2492	0.0577	0.7429	0.3455				
N	2460	3323	1551	2852				
	SAH	**Low malaise**	**MH problems**	**BMI**	**Chol ratio**	**Trig**	**CRP**	**Fib**
Hausman test	0.04	1.92	0.05	1.12	0.11	0.56	0.59	0.15
p-value	0.8328	0.1654	0.8307	0.2895	0.7433	0.4561	0.443	0.6978
N	2875	2854	2787	2759	2327	2333	2302	2295
SMC sample								
	School asp	**Job asp**	**Life sat**	**Self eff**	**Job posit**	**Crime**	**Drugs**	
Hausman test	0.29	0.14	0	0.02	0.29	0.34	0.45	
p-value	0.5916	0.7076	0.9953	0.8933	0.5919	0.559	0.5003	
N	4363	4473	3378	3330	3384	3547	3549	
	Income 33	**Employed 33**	**Income 50**	**Employed 50**				
Hausman test	3.66	0.01	1.8	1.16				
p-value	0.0559	0.9249	0.1797	0.2805				
N	2596	3595	1592	3034				
	SAH	**Low malaise**	**MH problems**	**BMI**	**Chol ratio**	**Trig**	**CRP**	**Fib**
Hausman test	0.01	1.67	0	0.16	0.45	0.17	0.14	0.17
p-value	0.9372	0.196	0.9461	0.6847	0.5037	0.6776	0.7071	0.6797
N	2597	3028	3003	2968	2483	2523	2490	2485

Table A2: 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
Background 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
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
Family socioeconomic 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
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 A3: Summary statistics of primary school characteristics by type of secondary school attended. Source: NCDS wave 3.

	Grammar				**Comprehensive**				**Secondary modern**			
	Mean	s.d.	Min	Max	Mean	s.d.	Min	Max	Mean	s.d.	Min	Max
Whether ability streaming in prim sch	0.38	0.49	0	1	0.34	0.47	0	1	0.37	0.48	0	1
Ability streaming age 11 - High track	0.33	0.47	0	1	0.12	0.33	0	1	0.09	0.28	0	1
Ability streaming age 11 - Avg track	0.04	0.20	0	1	0.11	0.31	0	1	0.13	0.34	0	1
Ability streaming age 11 - Low track	0.00	0.04	0	1	0.11	0.31	0	1	0.14	0.35	0	1
Class size	35.97	6.90	6	79	34.87	6.92	2	90	34.37	7.04	5	85
School size	331.32	140.96	17	987	328.41	138.70	10	999	324.06	141.43	18	876
No. regular teachers in school	2.62	1.63	1	9	2.45	1.52	1	9	2.44	1.49	1	9
No. full time teachers in schoo;	9.63	4.86	0	52	9.58	4.41	0	72	9.34	4.21	0	30
Perc. teachers with <1y experience	0.14	0.14	0	1	0.15	0.15	0	1	0.15	0.15	0	1
Perc. teachers with 1-2y experience	0.15	0.14	0	1	0.16	0.15	0	1	0.16	0.16	0	1
Perc. teachers with 3-5y experience	0.15	0.15	0	1	0.15	0.15	0	1	0.16	0.15	0	1
Perc. teachers with 6-10y experience	0.09	0.11	0	1	0.08	0.11	0	1	0.09	0.12	0	1
Perc. teachers with >10y experience	0.12	0.15	0	1	0.12	0.15	0	1	0.13	0.15	0	1
Observations	1151				5331				2289			

Table A4: Descriptive statistics of covariates by sample of estimation

	Dropped	**Age 16**	**Age 33**	**Age 33 wage**	**Age 42**	**Age 45**	**Age 50**	**Age 50 wage**
At birth								
Mother's age	27.44	27.50	27.51	27.57	27.51	27.51	27.58	27.55
Married mother	0.95	0.97	0.97	0.97	0.97	0.97	0.97	0.97
Husband SES	2.89	2.96	2.99	2.98	3.00	2.98	2.99	2.98
Mother's school	0.20	0.27	0.28	0.27	0.29	0.28	0.29	0.28
Abnormalities pregnancy	0.28	0.27	0.27	0.27	0.26	0.26	0.26	0.26
Pregnancy smoking	1.58	1.51	1.51	1.52	1.50	1.52	1.51	1.52
First born	0.26	0.35	0.35	0.34	0.36	0.36	0.36	0.34
Childhood								
Two or more siblings	-	0.67	0.66	0.66	0.66	0.66	0.65	0.66
No father figure	-	0.04	0.04	0.03	0.04	0.04	0.03	0.04
Child morbidity index	-	0.06	0.06	0.07	0.06	0.06	0.06	0.06
Chronic condition in the family	-	0.16	0.16	0.17	0.16	0.16	0.16	0.16
Cognitive skills	-	0.64	0.65	0.65	0.65	0.65	0.65	0.66
Non-cognitive skills	-	0.89	0.90	0.90	0.90	0.90	0.90	0.91
School type								
Grammar	-	0.13	0.14	0.15	0.14	0.14	0.15	0.17
Secondary modern	-	0.24	0.24	0.23	0.24	0.24	0.23	0.23
Comprehensive	-	0.54	0.54	0.55	0.54	0.54	0.54	0.53
Observations	12375	5878	4377	3438	3269	4603	4010	2150

Table A5: Models for wellbeing outcomes, OLS or non-linear, depending on outcome, using the entropy balanced grammar and comprehensive sample

	School	**Work**	**Life sat.**	**Self-eff.**	**Job positiv.**	**Crime**	**Drugs**
Grammar	0.3411**	0.5097*	-0.1538	-0.2335+	-0.0573	-0.1754	0.0690
	(0.1079)	(0.2396)	(0.1424)	(0.1250)	(0.1214)	(0.2261)	(0.2217)
Cogn skills qtile=2	0.0297	0.0777	-0.1213	0.0027	-0.0914	-0.0849	0.1711
	(0.0839)	(0.1772)	(0.1100)	(0.0971)	(0.0938)	(0.1766)	(0.1735)
Cogn skills qtile=3	0.1176	0.1936	-0.1479	-0.0775	0.0039	0.0373	-0.0322
	(0.0838)	(0.1822)	(0.1095)	(0.0968)	(0.0936)	(0.1724)	(0.1771)
Cogn skills qtile=4	0.1317	-0.0373	-0.0439	0.0802	0.0339	0.0119	0.2838
	(0.0863)	(0.1811)	(0.1169)	(0.1028)	(0.0993)	(0.1834)	(0.1761)
Grammar × CS qtile=2	-0.1203	-0.5442*	-0.0774	0.1984	0.1989	0.2929	-0.0698
	(0.1180)	(0.2604)	(0.1568)	(0.1376)	(0.1336)	(0.2532)	(0.2464)
Grammar × CS qtile=3	-0.2490*	-0.3348	0.1296	0.3217*	0.0806	0.1571	0.0534
	(0.1180)	(0.2765)	(0.1542)	(0.1357)	(0.1312)	(0.2467)	(0.2465)
Grammar × CS qtile=4	-0.1204	-0.1312	-0.0406	0.2061	0.0555	0.0890	-0.3891
	(0.1183)	(0.2728)	(0.1573)	(0.1381)	(0.1336)	(0.2568)	(0.2472)
Non-cogn skills qtile=2	0.1309+	0.2634	0.1077	0.1885*	0.0489	0.0010	-0.1558
	(0.0774)	(0.1653)	(0.1043)	(0.0920)	(0.0889)	(0.1629)	(0.1570)
Non-cogn skills qtile=3	0.1209	0.3480+	0.2331+	0.2405*	0.0686	0.0281	-0.2208
	(0.0911)	(0.2039)	(0.1208)	(0.1061)	(0.1023)	(0.1919)	(0.1868)
Non-cogn skills qtile=4	0.1245	0.1001	0.1264	0.1959*	-0.0172	-0.0299	-0.2213

	(0.0823)	(0.1710)	(0.1089)	(0.0957)	(0.0926)	(0.1771)	(0.1686)
Grammar × NCS qtile=2	-0.0818	-0.4060+	0.1045	0.0321	-0.1023	-0.1139	-0.0183
	(0.1100)	(0.2406)	(0.1461)	(0.1278)	(0.1241)	(0.2319)	(0.2221)
Grammar × NCS qtile=3	0.0167	-0.4107	-0.0177	-0.0763	-0.0098	-0.0306	0.1573
	(0.1295)	(0.2932)	(0.1711)	(0.1499)	(0.1454)	(0.2766)	(0.2634)
Grammar × NCS qtile=4	0.0050	0.2384	-0.0470	0.0033	0.0197	0.0687	-0.1393
	(0.1150)	(0.2696)	(0.1523)	(0.1331)	(0.1292)	(0.2445)	(0.2407)
Relative cogn. ability	1.4356***	1.1997***	-0.0573	0.3746*	0.8451***	-0.3330	0.5679+
	(0.1559)	(0.3303)	(0.2086)	(0.1828)	(0.1774)	(0.3328)	(0.3336)
Observations	4044	4156	3131	3083	3145	3277	3279
F	6.1004		1.0772	1.4757	3.2425		
chi2		60.6809				67.7050	69.0129

Standard errors in parentheses. All controls included. For probit models, raw coefficients are displayed.
+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table A6: Models for labour outcomes, OLS or non-linear, depending on outcome, using the entropy balanced grammar and comprehensive sample

	Log hourly wage 33	**Employed at 33**	**Log hourly wage 50**	**Employed at 50**
Grammar	0.1068	0.6594+	0.1289	1.0566*
	(0.1402)	(0.3739)	(0.1868)	(0.4415)
Cogn rank qtile=2	0.1282	0.0166	0.1680	0.0693
	(0.0807)	(0.1792)	(0.1111)	(0.2194)
Cogn rank qtile=3	0.1274	0.0420	0.1270	-0.0820
	(0.0791)	(0.1767)	(0.1075)	(0.2070)

Cogn rank qtile=4	0.3154***	0.2583	0.3310**	0.1237
	(0.0796)	(0.1878)	(0.1077)	(0.2180)
Grammar × CR qtile=2	0.0070	-0.0312	-0.1271	0.2050
	(0.1122)	(0.2580)	(0.1483)	(0.3149)
Grammar × CR qtile=3	0.0733	-0.1311	0.0095	0.1170
	(0.1119)	(0.2605)	(0.1469)	(0.3026)
Grammar × CR qtile=4	-0.1481	-0.3118	-0.0363	-0.2156
	(0.1114)	(0.2713)	(0.1484)	(0.3096)
Non-cogn skills qtile=2	0.0376	0.0305	0.2186*	-0.0802
	(0.0741)	(0.1764)	(0.1030)	(0.1956)
Non-cogn skills qtile=3	0.0694	-0.1114	0.0618	0.1861
	(0.0870)	(0.1959)	(0.1127)	(0.2418)
Non-cogn skills qtile=4	0.1222	0.0165	0.2285*	0.1669
	(0.0784)	(0.1827)	(0.1085)	(0.2167)
Grammar × NCS qtile=2	0.0689	-0.1249	-0.1023	0.1173
	(0.1031)	(0.2611)	(0.1415)	(0.2950)
Grammar × NCS qtile=3	0.0452	-0.1265	0.0794	-0.2725
	(0.1232)	(0.2889)	(0.1573)	(0.3410)
Grammar × NCS qtile=4	-0.0919	0.0567	-0.1013	-0.1722
	(0.1094)	(0.2689)	(0.1507)	(0.3095)
Relative cogn. ability	0.6482***	0.2787	0.7931***	-0.0732
	(0.1508)	(0.3501)	(0.2031)	(0.4093)
Observations	2460	3323	1551	2852

F	4.2820		2.2982	
chi2		194.3598		60.9997

Standard errors in parentheses. All controls included. For probit models, raw coefficients are displayed.
+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table A7: Models for health outcomes, OLS or non-linear, depending on outcome, using the entropy balanced grammar and comprehensive sample

	High SAH	**Low malaise**	**Mental ill-health**	**BMI**	**Chol ratio**	**Triglycerides**	**CRP**	**Fib**
Grammar	-0.2076	-0.1361	-0.0420	-0.1369	0.0791	0.1475	0.0244	-0.1234
	(0.1983)	(0.2271)	(0.1335)	(0.1470)	(0.1493)	(0.1441)	(0.1357)	(0.1569)
Cogn skills qtile=2	-0.1701	0.0667	-0.0801	-0.0056	0.0072	0.0413	-0.0351	0.0027
	(0.1552)	(0.1818)	(0.1042)	(0.1149)	(0.1158)	(0.1124)	(0.1057)	(0.1218)
Cogn skills qtile=3	-0.1675	-0.2741	-0.0252	-0.0544	0.0570	-0.0182	-0.1323	-0.0972
	(0.1535)	(0.1731)	(0.1045)	(0.1151)	(0.1153)	(0.1120)	(0.1056)	(0.1215)
Cogn skills qtile=4	0.1433	0.0407	-0.0175	-0.1882	-0.1428	-0.1365	-0.0984	-0.0828
	(0.1614)	(0.1872)	(0.1071)	(0.1177)	(0.1178)	(0.1144)	(0.1068)	(0.1229)
Grammar × CS qtile=2	0.3497	0.0054	0.0909	0.0691	-0.0746	-0.0742	0.0620	0.1542
	(0.2173)	(0.2518)	(0.1469)	(0.1623)	(0.1642)	(0.1592)	(0.1499)	(0.1728)
Grammar × CS qtile=3	0.3074	0.4924*	0.0700	0.0893	-0.1013	0.0253	0.1041	0.1405
	(0.2125)	(0.2435)	(0.1448)	(0.1593)	(0.1611)	(0.1562)	(0.1478)	(0.1702)
Grammar × CS qtile=4	0.1876	0.1344	-0.0258	-0.0817	0.0305	0.0486	0.0655	0.1835
	(0.2186)	(0.2518)	(0.1459)	(0.1606)	(0.1603)	(0.1554)	(0.1463)	(0.1683)
Non-cogn skills qtile=2	0.1058	0.0274	-0.0566	-0.0883	-0.0879	0.0327	-0.0224	-0.1048

	(0.1446)	(0.1665)	(0.0976)	(0.1075)	(0.1077)	(0.1047)	(0.0982)	(0.1130)
Non-cogn skills qtile=3	0.0927	0.0713	-0.1433	-0.1162	-0.1205	-0.0420	0.0233	-0.0631
	(0.1663)	(0.1934)	(0.1141)	(0.1255)	(0.1264)	(0.1228)	(0.1163)	(0.1337)
Non-cogn skills qtile=4	0.1126	0.0417	-0.0542	-0.0377	-0.0506	0.0143	-0.0116	-0.1094
	(0.1514)	(0.1745)	(0.1027)	(0.1130)	(0.1130)	(0.1097)	(0.1025)	(0.1179)
Grammar × NCS qtile=2	0.0038	0.1191	-0.0213	0.0557	0.0158	-0.2332	-0.0881	-0.0050
	(0.2036)	(0.2363)	(0.1376)	(0.1517)	(0.1549)	(0.1502)	(0.1416)	(0.1630)
Grammar × NCS qtile=3	-0.0410	0.1588	0.1231	0.0992	-0.0371	-0.1865	-0.2250	0.0140
	(0.2333)	(0.2747)	(0.1601)	(0.1762)	(0.1801)	(0.1745)	(0.1645)	(0.1896)
Grammar × NCS qtile=4	-0.0569	-0.0797	0.0190	-0.0776	-0.0197	-0.1953	-0.0553	0.0607
	(0.2109)	(0.2404)	(0.1418)	(0.1560)	(0.1572)	(0.1522)	(0.1434)	(0.1655)
Relative cogn. ability	0.2486	0.4524	0.1993	0.0178	-0.2583	-0.1284	-0.4311*	-0.4286+
	(0.2898)	(0.3306)	(0.1969)	(0.2168)	(0.2204)	(0.2133)	(0.2013)	(0.2327)
Observations	2875	2854	2787	2759	2327	2333	2302	2295
F			1.1660	1.5968	5.6317	4.3801	0.6766	1.4088
chi2	54.0547	67.1764						

Standard errors in parentheses. All controls included. For probit models, raw coefficients are displayed.
+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table A8: Models for wellbeing outcomes, OLS or non-linear, depending on outcome, using the entropy balanced secondary modern and comprehensive.

	School	**Work**	**Life sat.**	**Self-eff.**	**Job positiv.**	**Crime**	**Drugs**
[1em] Secondary modern	0.0153	0.1079	0.0054	0.1640	0.0661	-0.0860	-0.0437

	(0.0798)	(0.1288)	(0.1170)	(0.1126)	(0.1117)	(0.1619)	(0.1685)
Cogn skills qtile=2	0.0032	0.0818	0.0339	0.1632+	0.2286*	0.0710	0.0766
	(0.0660)	(0.1101)	(0.0957)	(0.0918)	(0.0912)	(0.1371)	(0.1428)
Cogn skills qtile=3	-0.0310	-0.0373	0.0443	0.1217	0.2640**	0.0651	0.0296
	(0.0672)	(0.1139)	(0.0970)	(0.0933)	(0.0928)	(0.1412)	(0.1473)
Cogn skills qtile=4	0.0641	0.0069	-0.0022	0.1699+	0.1400	0.0165	0.1958
	(0.0694)	(0.1202)	(0.1014)	(0.0981)	(0.0970)	(0.1473)	(0.1489)
Secondary modern × CS qtile=2	0.0436	0.0678	0.1205	0.0430	0.0140	-0.1120	0.1051
	(0.0905)	(0.1521)	(0.1314)	(0.1260)	(0.1251)	(0.1923)	(0.1979)
Secondary modern × CS qtile=3	-0.0085	0.1466	0.0235	-0.0239	-0.1816	-0.0092	0.1550
	(0.0894)	(0.1528)	(0.1302)	(0.1253)	(0.1244)	(0.1898)	(0.1968)
Secondary modern × CS qtile=4	-0.0917	-0.0421	0.1559	-0.0121	0.0112	0.1402	-0.1276
	(0.0903)	(0.1574)	(0.1317)	(0.1270)	(0.1257)	(0.1916)	(0.1988)
Non-cogn skills qtile=2	0.0152	0.2866**	0.0957	0.0796	0.0349	-0.1970	-0.1373
	(0.0617)	(0.1040)	(0.0893)	(0.0864)	(0.0853)	(0.1258)	(0.1274)
Non-cogn skills qtile=3	0.0162	0.1711+	0.1558+	0.1083	0.0399	-0.1724	-0.2055
	(0.0610)	(0.1036)	(0.0884)	(0.0848)	(0.0838)	(0.1253)	(0.1271)
Non-cogn skills qtile=4	0.1019	0.2237+	0.2287*	0.1850+	0.1106	-0.1027	-0.3688*
	(0.0701)	(0.1232)	(0.1009)	(0.0967)	(0.0957)	(0.1467)	(0.1549)
Secondary modern × NCS qtile=2	0.0809	-0.2166	-0.2373+	-0.1823	-0.0396	0.0217	-0.0025
	(0.0853)	(0.1454)	(0.1247)	(0.1201)	(0.1188)	(0.1773)	(0.1778)
Secondary modern × NCS qtile=3	0.1475+	-0.0876	-0.0613	-0.0852	0.0692	0.0510	-0.0101

	(0.0850)	(0.1473)	(0.1234)	(0.1188)	(0.1172)	(0.1756)	(0.1783)
Secondary modern × NCS qtile=4	0.1831+	-0.0494	-0.0326	-0.0649	-0.0338	-0.1474	-0.1384
	(0.0960)	(0.1722)	(0.1383)	(0.1323)	(0.1313)	(0.2039)	(0.2170)
Relative cogn. ability	1.2655***	1.5738***	-0.1825	0.2573*	0.6309***	-0.1372	0.2189
	(0.0929)	(0.1751)	(0.1356)	(0.1299)	(0.1289)	(0.1986)	(0.2029)
Observations	4665	4818	3588	3535	3597	3777	3779
F	15.6073		1.9836	2.4496	7.1127		
chi2		285.6332				214.3746	87.1143

Standard errors in parentheses. All controls included.
+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table A9: Models for labour outcomes, OLS or non-linear, depending on outcome, using the entropy balanced secondary modern and comprehensive sample

	Log hourly wage 33	**Employed at 33**	**Log hourly wage 50**	**Employed at 50**
Secmodern	-0.0003	0.0122	0.2793*	-0.2912
	(0.0791)	(0.1919)	(0.1358)	(0.2126)
Cogn rank qtile=2	0.0400	0.1658	0.0848	0.2869+
	(0.0554)	(0.1280)	(0.0951)	(0.1503)
Cogn rank qtile=3	0.1413*	0.1180	0.2071*	0.3274*
	(0.0557)	(0.1313)	(0.0934)	(0.1585)
Cogn rank qtile=4	0.2070***	0.2013	0.2781**	0.3968*
	(0.0558)	(0.1344)	(0.0919)	(0.1591)
Secmodern × CR qtile=2	0.1002	0.1754	-0.0632	0.0444

	(0.0784)	(0.1901)	(0.1353)	(0.2137)
Secmodern × CR qtile=3	0.0830	0.1294	-0.0624	0.0343
	(0.0778)	(0.1861)	(0.1321)	(0.2165)
Secmodern × CR qtile=4	0.0429	0.0365	-0.0987	0.1123
	(0.0785)	(0.1934)	(0.1302)	(0.2218)
Non-cogn skills qtile=2	0.0285	0.0592	0.0615	0.0542
	(0.0511)	(0.1239)	(0.0893)	(0.1448)
Non-cogn skills qtile=3	0.0037	0.0909	0.0401	0.1418
	(0.0507)	(0.1233)	(0.0851)	(0.1455)
Non-cogn skills qtile=4	0.0920	0.1053	0.0967	0.2952+
	(0.0584)	(0.1408)	(0.0967)	(0.1757)
Secmodern × NCS qtile=2	-0.0915	0.0669	-0.1210	0.0943
	(0.0730)	(0.1807)	(0.1240)	(0.2033)
Secmodern × NCS qtile=3	0.0506	0.0276	-0.0984	0.1423
	(0.0726)	(0.1827)	(0.1197)	(0.2077)
Secmodern × NCS qtile=4	0.0267	0.0821	-0.1351	-0.0258
	(0.0820)	(0.2040)	(0.1335)	(0.2422)
Relative cogn. ability	0.3664***	0.3014	0.4569***	0.4870*
	(0.0779)	(0.1982)	(0.1264)	(0.2338)
Observations	2766	3821	1689	3230
F	11.3550		2.8229	
chi2		277.5268		112.3251

Standard errors in parentheses. All controls included. For probit models, raw coefficients are displayed.

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

Table A10: Models for health outcomes, OLS or non-linear, depending on outcome, using the entropy balanced secondary modern and comprehensive sample

	High SAH	Low malaise	Mental ill-health	BMI	Chol ratio	Triglycerides	CRP	Fib
Secondary modern	-0.0060	0.0091	0.0030	-0.0111	-0.1297	-0.1745	0.1137	0.0418
	(0.1594)	(0.1683)	(0.1235)	(0.1257)	(0.1299)	(0.1305)	(0.1503)	(0.1349
Cogn skills qtile=2	0.0143	-0.1326	0.0015	-0.0580	-0.1104	-0.0522	0.0548	-0.1179
	(0.1254)	(0.1355)	(0.1004)	(0.1021)	(0.1045)	(0.1050)	(0.1211)	(0.1087
Cogn skills qtile=3	0.1549	-0.0781	-0.0069	-0.0663	-0.1853+	-0.1564	-0.0301	-0.0399
	(0.1298)	(0.1421)	(0.1016)	(0.1032)	(0.1060)	(0.1064)	(0.1225)	(0.1100
Cogn skills qtile=4	-0.0399	-0.0671	-0.0435	-0.0477	-0.1219	-0.1062	-0.0141	-0.0403
	(0.1333)	(0.1463)	(0.1054)	(0.1072)	(0.1112)	(0.1118)	(0.1287)	(0.1154
Secondary modern × CS qtile=2	0.0588	0.1569	-0.1025	0.1798	0.0792	0.0444	-0.0512	0.0657
	(0.1759)	(0.1881)	(0.1379)	(0.1406)	(0.1440)	(0.1449)	(0.1670)	(0.1498
Secondary modern × CS qtile=3	-0.0923	0.2473	-0.1584	0.1024	0.2085	0.2422+	-0.0393	0.0635
	(0.1739)	(0.1896)	(0.1349)	(0.1374)	(0.1412)	(0.1419)	(0.1626)	(0.1459
Secondary modern × CS qtile=4	0.1829	0.2894	-0.1312	0.0793	0.1984	0.1815	-0.1547	-0.0093
	(0.1759)	(0.1937)	(0.1384)	(0.1409)	(0.1458)	(0.1466)	(0.1689)	(0.1515
Non-cogn skills qtile=2	0.1031	0.2443+	-0.0832	0.0507	-0.0399	-0.0930	0.1096	0.1403
	(0.1197)	(0.1285)	(0.0951)	(0.0967)	(0.1005)	(0.1010)	(0.1165)	(0.1045
Non-cogn skills qtile=3	0.2409*	0.3203*	-0.1636+	0.0479	-0.0583	-0.1012	0.0835	0.1365

	(0.1165)	(0.1257)	(0.0925)	(0.0939)	(0.0969)	(0.0972)	(0.1126)	(0.1011
Non-cogn skills qtile=4	0.2830*	0.3953**	-0.2589*	0.1405	-0.0295	-0.0127	0.2044	0.0840
	(0.1331)	(0.1466)	(0.1066)	(0.1081)	(0.1108)	(0.1111)	(0.1282)	(0.1152
Secondary modern × NCS qtile=2	-0.0820	-0.2457	0.0561	-0.0250	0.1244	0.1137	-0.1115	-0.105
	(0.1664)	(0.1797)	(0.1308)	(0.1331)	(0.1383)	(0.1391)	(0.1603)	(0.1438
Secondary modern × NCS qtile=3	-0.0166	-0.2389	0.0726	0.0567	0.1538	0.1515	0.0234	-0.123
	(0.1633)	(0.1777)	(0.1279)	(0.1303)	(0.1348)	(0.1355)	(0.1571)	(0.1410
Secondary modern × NCS qtile=4	-0.0688	-0.1424	0.1772	-0.2748+	-0.0453	-0.0203	-0.2474	-0.172
	(0.1829)	(0.2038)	(0.1432)	(0.1452)	(0.1493)	(0.1499)	(0.1728)	(0.1551
Relative cogn. ability	0.4071*	0.4304*	-0.1313	-0.1405	-0.1248	-0.1296	-0.1640	-0.4371
	(0.1802)	(0.2017)	(0.1399)	(0.1424)	(0.1476)	(0.1488)	(0.1709)	(0.1532
Observations	3250	3224	3183	3145	2665	2669	2634	2629
F			2.4141	2.2009	5.0753	3.3191	1.5678	2.2067
chi2	111.4146	99.7268						

Standard errors in parentheses. All controls included. For probit models, raw coefficients are displayed.

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

Table A11: Models for wellbeing outcomes, distinguishing between comprehensives by origin. Base category is grammar (balanced grammar and comprehensive sample).

	School	**Work**	**Life sat.**	**Self-eff.**	**Job positiv.**	**Crime**	**Drugs**
Comprehensive (former grammar)	-0.1797**	-0.0209	0.1479+	0.0506	0.0475	0.0354	0.0033
	(0.0655)	(0.0214)	(0.0852)	(0.0753)	(0.0724)	(0.0301)	(0.0324)
Comprehensive (other)	-0.2040***	-0.0103	0.1241*	0.0547	-0.0163	0.0060	0.0129
	(0.0460)	(0.0152)	(0.0611)	(0.0537)	(0.0520)	(0.0220)	(0.0232)
Observations	4044	4156	3131	3083	3145	3277	3279
F	7.9541		1.1901	1.6806	4.0877		
chi2		52.50				64.39	66.08

Standard errors in parentheses. All controls included. For probit models, marginal effects are displayed.

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

Table A12: Models for labour outcomes, distinguishing between comprehensives by origin. Base category is grammar (balanced grammar and comprehensive sample).

	Log hourly wage 33	**Employed at 33**	**Log hourly wage 50**	**Employed at 50**
Comprehensive (former grammar)	-0.0609	0.0052	-0.0711	0.0144
	(0.0609)	(0.0301)	(0.0799)	(0.0279)
Comprehensive (other)	-0.0579	-0.0462*	-0.0950	-0.0309+
	(0.0436)	(0.0213)	(0.0582)	(0.0182)
Observations	2460	3323	1551	2852
F	5.1292		2.6448	
chi2		191.98		50.09

Standard errors in parentheses. All controls included. For probit models, marginal effects are displayed.

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

Table A13: Models for health outcomes, distinguishing between comprehensives by origin. Base category is grammar (balanced grammar and comprehensive sample).

	High SAH	**Low malaise**	**Mental ill-health**	**BMI**	**Chol ratio**	**Trig**	**CRP**	**Fib**
Comprehensive (former grammar)	-0.0023	0.0104	0.0014	0.0419	-0.1452+	-0.0413	-0.0415	-0.1660+
	(0.0437)	(0.0356)	(0.0789)	(0.0871)	(0.0867)	(0.0843)	(0.0785)	(0.0901)
Comprehensive (other)	0.0085	-0.0272	-0.0176	0.1310*	0.0045	0.0263	0.0130	0.0507
	(0.0313)	(0.0251)	(0.0574)	(0.0637)	(0.0636)	(0.0618)	(0.0585)	(0.0671)
Observations	2875	2854	2787	2759	2327	2333	2302	2295
F			1.4210	2.0588	7.1200	5.5134	0.5376	1.6736
chi2	52.9100	57.4897						

Standard errors in parentheses. All controls included. For probit models, marginal effects are displayed.

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

Table A14: Models for wellbeing outcomes, distinguishing between comprehensives by origin. Base category is secondary modern (balanced secondary modern and comprehensive sample).

	School	**Work**	**Life sat.**	**Self-eff.**	**Job positiv.**	**Crime**	**Drugs**
Comprehensive (former sec modern)	-0.1179*	-0.0386+	0.0458	-0.0551	-0.0091	0.0051	-0.0192
	(0.0477)	(0.0228)	(0.0697)	(0.0668)	(0.0663)	(0.0243)	(0.0239)
Comprehensive (other)	-0.0915**	-0.0094	-0.0084	-0.0886+	-0.0344	0.0266	0.0166
	(0.0346)	(0.0171)	(0.0503)	(0.0483)	(0.0479)	(0.0176)	(0.0168)
Observations	4665	4818	3588	3535	3597	3777	3779
F	20.4740		2.1065	2.9852	9.1645		
chi2		276.89				207.54	86.04

Standard errors in parentheses. All controls included. For probit models, marginal effects are displayed.

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

Table A15: Models for labour outcomes, distinguishing between comprehensives by origin. Base category is secondary modern (balanced secondary modern and comprehensive sample).

	Log hourly wage 33	**Employed at 33**	**Log hourly wage 50**	**Employed at 50**
Comprehensive (former sec modern)	-0.0214	-0.0340	-0.0929	0.0257
	(0.0392)	(0.0244)	(0.0621)	(0.0248)
Comprehensive (other)	-0.0589*	-0.0343+	-0.0820+	0.0173
	(0.0290)	(0.0180)	(0.0464)	(0.0178)
Observations	2766	3821	1689	3230
F	14.4193		3.4878	
chi2		267.75		109.31

Standard errors in parentheses. All controls included. For probit models, marginal effects are displayed.

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

Table A16: Models for health outcomes, distinguishing between comprehensives by origin. Base category is secondary modern (balanced secondary modern and comprehensive sample).

	High SAH	**Low malaise**	**Mental ill-health**	**BMI**	**Chol ratio**	**Trig**	**CRP**	**Fib**
Comprehensive (former sec modern)	0.0122	-0.0281	-0.0350	-0.1076	-0.1016	0.0001	-0.0847	-0.0887
	(0.0348)	(0.0294)	(0.0714)	(0.0726)	(0.0748)	(0.0749)	(0.0867)	(0.0779)
Comprehensive (other)	0.0009	0.0032	0.0547	-0.0092	-0.0509	-0.0326	0.0649	0.0750
	(0.0254)	(0.0219)	(0.0520)	(0.0531)	(0.0545)	(0.0548)	(0.0632)	(0.0566)
Observations	3250	3224	3183	3145	2665	2669	2634	2629
F			3.0059	2.3316	6.1448	4.0868	1.8526	2.4622
chi2	109.3500	97.7415						

Standard errors in parentheses. All controls included. For probit models, marginal effects are displayed.

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

Table A17: First stage for grammar and secondary modern attendance, using % pupils going to comprehensive schools in individual's LEA as an IV

	(1) Grammar	(2) Grammar	(3) Sec modern	(4) 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 A18: IV models for wellbeing outcomes (balanced grammar and comprehensive sample).

	School	**Work**	**Life sat.**	**Self-eff.**	**Job positiv.**	**Crime**	**Drugs**
Grammar	0.1088	0.0028	-0.1621	-0.0348	-0.0137	-0.0324	0.0725
	(0.1002)	(0.0337)	(0.1297)	(0.1137)	(0.1105)	(0.0467)	(0.0493)
Cognitive skills	0.1597	-0.0783	-0.0527	0.5896*	0.2403	0.0750	-0.0178
	(0.2216)	(0.0749)	(0.2950)	(0.2590)	(0.2502)	(0.1071)	(0.1130)
Non-cognitive skills	0.7498**	0.0753	1.1859**	1.1129***	0.0421	-0.0502	-0.3342*
	(0.2783)	(0.0936)	(0.3800)	(0.3331)	(0.3212)	(0.1342)	(0.1416)
Relative cogn. ability	1.4351***	0.2032***	-0.1181	0.3586+	0.8331***	-0.0791	0.1563+
	(0.1576)	(0.0533)	(0.2103)	(0.1844)	(0.1785)	(0.0774)	(0.0817)
Female	0.0079	0.0114	0.0653	-0.0690	-0.3798***	-0.1126***	-0.0652**
	(0.0427)	(0.0144)	(0.0565)	(0.0495)	(0.0478)	(0.0207)	(0.0219)
Mother's interest	0.1037***	0.0016	0.0563	-0.0110	0.0639+	0.0239+	0.0214
	(0.0285)	(0.0096)	(0.0383)	(0.0337)	(0.0329)	(0.0144)	(0.0152)
Father's SES	0.0267	0.0224	0.0813	-0.0356	0.0595	-0.0859**	-0.0032
	(0.0632)	(0.0213)	(0.0836)	(0.0734)	(0.0713)	(0.0303)	(0.0320)
Observations	4044	4156	3131	3083	3145	3277	3279
F	7.6007	1.3948	1.1180	1.6972	4.1800	1.7519	1.7779

Standard errors in parentheses. Father's SES is a dummy for high/middle-high SES. All controls included.

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

Table A19: IV models for labour outcomes (balanced grammar and comprehensive sample).

	Log hourly wage 33	**Employed at 33**	**Log hourly wage 50**	**Employed at 50**
Grammar	0.1565+	0.1063*	0.0548	0.0455
	(0.0941)	(0.0465)	(0.1142)	(0.0392)
Cognitive skills	0.0017	0.0844	-0.0215	0.0702
	(0.2081)	(0.1063)	(0.2750)	(0.0910)
Non-cognitive skills	0.4218	0.0284	0.7626*	0.0984
	(0.2701)	(0.1352)	(0.3650)	(0.1178)
Relative cogn. ability	0.6644***	0.0789	0.7868***	-0.0038
	(0.1525)	(0.0753)	(0.2059)	(0.0661)
Female	-0.4294***	-0.2407***	-0.2216***	-0.0624***
	(0.0402)	(0.0205)	(0.0542)	(0.0178)
Mother's interest	-0.0070	0.0033	0.0552	0.0059
	(0.0276)	(0.0139)	(0.0375)	(0.0122)
Father's SES	0.0279	0.0400	-0.0539	0.0216
	(0.0598)	(0.0301)	(0.0806)	(0.0262)
Observations	2460	3323	1551	2852
F	5.2506	4.8360	2.6444	1.2194

Standard errors in parentheses. Father's SES is a dummy for high/middle-high SES. All controls included.

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

Table A20: IV models for health outcomes (balanced grammar and comprehensive sample).

	High SAH	**Low malaise**	**Mental ill-health**	**BMI**	**Chol ratio**	**Triglycerides**	**CRP**	**Fibrinogen**
Grammar	0.0076	0.0802	0.0256	-0.2357+	-0.0021	-0.0957	-0.0834	-0.0376
	(0.0663)	(0.0534)	(0.1208)	(0.1351)	(0.1355)	(0.1319)	(0.1249)	(0.1432)
Cognitive skills	0.2238	0.0723	-0.1827	-0.6264*	-0.5397+	-0.5752*	-0.1335	-0.0611
	(0.1526)	(0.1223)	(0.2707)	(0.2995)	(0.2984)	(0.2892)	(0.2717)	(0.3124)
Non-cognitive skills	0.2247	0.1004	-0.6175+	-0.4060	-0.3511	-0.2466	-0.2144	-0.5380
	(0.1965)	(0.1586)	(0.3479)	(0.3855)	(0.3857)	(0.3749)	(0.3535)	(0.4066)
Relative cogn. ability	0.0970	0.1216	0.2424	-0.0065	-0.2198	-0.0850	-0.4713*	-0.4247+
	(0.1107)	(0.0890)	(0.1985)	(0.2201)	(0.2214)	(0.2146)	(0.2028)	(0.2339)
Female	0.0176	-0.0986***	0.2171***	-0.2399***	-0.8503***	-0.6731***	0.0089	0.2355***
	(0.0295)	(0.0237)	(0.0527)	(0.0583)	(0.0587)	(0.0571)	(0.0537)	(0.0618)
Mother's interest	0.0056	0.0080	0.0152	-0.0711+	-0.0277	-0.0574	-0.0211	-0.0692
	(0.0205)	(0.0165)	(0.0378)	(0.0420)	(0.0425)	(0.0413)	(0.0390)	(0.0448)
Father's SES	0.0137	-0.0516	-0.0182	-0.1073	-0.0134	-0.1212	-0.0218	-0.0360
	(0.0439)	(0.0353)	(0.0773)	(0.0858)	(0.0882)	(0.0857)	(0.0812)	(0.0933)
Observations	2875	2854	2787	2759	2327	2333	2302	2295
F	1.3807	1.5620	1.4590	2.0726	7.1932	5.6424	0.5509	1.5584

Standard errors in parentheses. Father's SES is a dummy for high/middle-high SES. All controls included.

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

Table A21: IV models for wellbeing outcomes (balanced secondary modern and comprehensive sample).

	School	**Work**	**Life sat.**	**Self-eff.**	**Job positiv.**	**Crime**	**Drugs**
Secondary modern	0.1243+	0.0538	0.0002	0.0646	-0.0380	-0.0249	0.0064
	(0.0682)	(0.0342)	(0.1011)	(0.0975)	(0.0974)	(0.0372)	(0.0355)
Cognitive skills	-0.0869	0.0652	0.2020	0.3836*	0.2550	0.0886	0.1088
	(0.1318)	(0.0655)	(0.1969)	(0.1888)	(0.1875)	(0.0701)	(0.0667)
Non-cognitive skills	0.5864***	0.1638*	0.5849**	0.3927*	0.4669*	-0.1398*	-0.2623***
	(0.1342)	(0.0663)	(0.1976)	(0.1899)	(0.1886)	(0.0690)	(0.0656)
Relative cogn. ability	1.2833***	0.3920***	-0.2024	0.2253+	0.6089***	-0.0538	0.0336
	(0.0945)	(0.0470)	(0.1389)	(0.1330)	(0.1321)	(0.0496)	(0.0472)
Female	0.0143	-0.0070	0.1551***	-0.0186	-0.5753***	-0.2023***	-0.0828***
	(0.0311)	(0.0154)	(0.0452)	(0.0433)	(0.0430)	(0.0161)	(0.0154)
Mother's interest	0.1074***	0.0149+	0.0085	0.0221	0.0256	0.0059	0.0099
	(0.0159)	(0.0079)	(0.0231)	(0.0222)	(0.0220)	(0.0082)	(0.0078)
Father's SES	0.1093**	0.0345+	0.0459	0.0818	0.0498	-0.0149	-0.0215
	(0.0371)	(0.0183)	(0.0542)	(0.0521)	(0.0517)	(0.0194)	(0.0185)
Observations	4665	4818	3588	3535	3597	3777	3779
F	20.8419	7.7455	2.1480	2.9785	9.3945	5.7980	2.3666

Standard errors in parentheses. Father's SES is a dummy for high/middle-high SES. All controls included.

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

Table A22: IV models for wellbeing outcomes (balanced secondary modern and comprehensive sample).

	Log hourly wage 33	**Employed at 33**	**Log hourly wage 50**	**Employed at 50**
Secondary modern	-0.0527	0.0272	-0.0157	-0.0551
	(0.0602)	(0.0373)	(0.0970)	(0.0379)
Cognitive skills	0.1824	0.0468	0.0429	0.1177
	(0.1135)	(0.0715)	(0.1861)	(0.0717)
Non-cognitive skills	0.2695*	0.1243+	-0.0325	0.2129**
	(0.1132)	(0.0719)	(0.1934)	(0.0687)
Relative cogn. ability	0.3659***	0.0505	0.4411***	0.1077*
	(0.0802)	(0.0506)	(0.1293)	(0.0502)
Female	-0.4783***	-0.2444***	-0.2681***	-0.0808***
	(0.0260)	(0.0165)	(0.0417)	(0.0163)
Mother's interest	0.0295*	0.0100	0.0415+	-0.0009
	(0.0134)	(0.0084)	(0.0213)	(0.0082)
Father's SES	0.0907**	0.0577**	0.0206	0.0225
	(0.0308)	(0.0198)	(0.0507)	(0.0197)
Observations	2766	3821	1689	3230
F	14.2238	7.1046	3.4053	3.0578

Standard errors in parentheses. Father's SES is a dummy for high/middle-high SES. All controls included.

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

Table A23: IV models for health outcomes (balanced secondary modern and comprehensive sample).

	High SAH	**Low malaise**	**Mental ill-health**	**BMI**	**Chol ratio**	**Triglycerides**	**CRP**	**Fibrinogen**
Secondary modern	-0.0051	-0.0281	-0.0549	-0.0184	0.1290	0.0874	-0.0004	-0.0198
	(0.0538)	(0.0463)	(0.1079)	(0.1115)	(0.1133)	(0.1143)	(0.1325)	(0.1186)
Cognitive skills	0.0588	0.0371	-0.1498	-0.0874	-0.0241	-0.0813	-0.2529	-0.0474
	(0.1016)	(0.0875)	(0.2049)	(0.2093)	(0.2158)	(0.2175)	(0.2494)	(0.2237)
Non-cognitive skills	0.3439***	0.3668***	-0.6086**	0.1203	-0.2716	-0.2172	0.1968	0.0893
	(0.0974)	(0.0839)	(0.2022)	(0.2055)	(0.2154)	(0.2165)	(0.2445)	(0.2193)
Relative cogn. ability	0.1527*	0.1246*	-0.1569	-0.1355	-0.1006	-0.0812	-0.1582	-0.4423**
	(0.0711)	(0.0612)	(0.1432)	(0.1462)	(0.1515)	(0.1526)	(0.1753)	(0.1572)
Female	-0.0326	-0.1105***	0.2567***	-0.1663***	-0.6612***	-0.5396***	0.1117+	0.2688***
	(0.0231)	(0.0198)	(0.0467)	(0.0477)	(0.0493)	(0.0495)	(0.0571)	(0.0513)
Mother's interest	0.0268*	-0.0173+	0.0166	-0.0798**	-0.0483+	-0.0065	-0.0113	-0.0372
	(0.0116)	(0.0100)	(0.0238)	(0.0243)	(0.0250)	(0.0251)	(0.0290)	(0.0260)
Father's SES	0.0235	-0.0065	0.0510	-0.1522**	-0.0021	-0.0216	-0.0301	-0.0112
	(0.0279)	(0.0241)	(0.0559)	(0.0571)	(0.0590)	(0.0594)	(0.0684)	(0.0613)
Observations	3250	3224	3183	3145	2665	2669	2634	2629
F	2.9665	2.6751	3.0451	2.3339	6.2781	4.1983	1.8196	2.4052

Standard errors in parentheses. Father's SES is a dummy for high/middle-high SES. All controls included.

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

Table A24: Regressions for age 16 maths scores and placebo regressions for age 11 maths scores (balanced grammar and comprehensive sample).

	Age 16 maths scores				Age 11 maths scores			
	OLS	OLS	2SLS	2SLS	OLS	OLS	2SLS	2SLS
Comprehensive	-0.0566***	-0.0655***	-0.0371+	-0.0493**	-0.0882***	-0.0859***	-0.1293***	-0.0940***
	(0.0072)	(0.0078)	(0.0202)	(0.0190)	(0.0071)	(0.0079)	(0.0194)	(0.0188)
Cognitive ability 11	0.9207***	0.8509***	0.9442***	0.8697***				
	(0.0323)	(0.0354)	(0.0395)	(0.0408)				
Cognitive ability 7					0.5804***	0.5187***	0.5776***	0.5192***
					(0.0356)	(0.0390)	(0.0368)	(0.0390)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	5282	4171	5282	4171	5282	4171	5282	4171
F	535.7024	31.7958	504.1154	29.8687	210.4844	12.5699	145.2826	9.9248

Standard errors in parentheses

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

Table A25: Regressions for age 16 BMI and placebo regressions for age 11 BMI (balanced grammar and comprehensive sample).

	Age 16 BMI				Age 11 BMI			
	OLS	OLS	2SLS	2SLS	OLS	OLS	2SLS	2SLS
Comprehensive	0.1237	0.0230	-0.0904	-0.0981	0.0459	-0.0555	-0.4638	-0.4920
	(0.1410)	(0.1512)	(0.3616)	(0.3622)	(0.1260)	(0.1382)	(0.3085)	(0.3208)
Cognitive ability 11	0.0097	0.0298	-0.0243	0.0114				
	(0.0807)	(0.0870)	(0.0966)	(0.1004)				
Cognitive ability 7					0.0098	0.0052	0.0183	0.0154
					(0.0789)	(0.0844)	(0.0795)	(0.0850)
Non-cognitive skills			-1.2618		-1.2663		-1.7265+	
-1.7670+								
		(0.9712)		(0.9719)		(0.9215)		(0.9258)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	3777	3294	3777	3294	3777	3294	3777	3294
F	0.3931	1.3578	0.0395	1.3575	0.0754	1.1916	1.1392	1.2425

Standard errors in parentheses

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

Table A26: Regressions for age 16 maths scores and placebo regressions for age 11 maths scores (balanced secondary modern and comprehensive sample).

	Age 16 maths scores				Age 11 maths scores			
	OLS	OLS	2SLS	2SLS	OLS	OLS	2SLS	2SLS
Comprehensive	0.0056	0.0055	0.0095	0.0120	-0.0453***	-0.0427***	-0.0550***	-0.0454***
	(0.0044)	(0.0050)	(0.0097)	(0.0112)	(0.0052)	(0.0059)	(0.0113)	(0.0131)
Cognitive ability 11	0.6601***	0.6309***	0.6592***	0.6337***				
	(0.0149)	(0.0181)	(0.0158)	(0.0186)				
Cognitive ability 7					0.7151***	0.6441***	0.7079***	0.6443***
					(0.0176)	(0.0204)	(0.0178)	(0.0204)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	6181	4865	6181	4865	6181	4865	6181	4865
F	987.4593	50.3331	929.0766	50.2900	865.2405	45.9691	801.4544	44.8148

Standard errors in parentheses

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

Table A27: Regression for BMI at age 16 and placebo regression for BMI age 11 (balanced secondary modern and comprehensive sample).

	(1) OLS	(2) OLS	(3) 2SLS	(4) 2SLS	(5) OLS	(6) OLS	(7) 2SLS	(8) 2SLS
Comprehensive	-0.0301	-0.1304	-0.0668	-0.1748	-0.1011	-0.1726	-0.1529	-0.2550
	(0.1203)	(0.1295)	(0.2501)	(0.2809)	(0.1015)	(0.1103)	(0.2087)	(0.2380)
Cognitive ability 11	0.0199	0.0641	0.0177	0.0614				
	(0.0514)	(0.0596)	(0.0531)	(0.0614)				
Cognitive skills 7					0.0763+	0.0606	0.0766+	0.0616
					(0.0425)	(0.0476)	(0.0425)	(0.0476)
Non-cognitive skills		-0.5617		-0.5565		0.2213		0.2183
		(0.5267)		(0.5275)		(0.4437)		(0.4438)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	4384	3816	4384	3816	4384	3816	4384	3816
F	0.1230	2.5103	0.1275	2.4924	2.0854	2.4015	1.8576	2.3644

Standard errors in parentheses

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

A.3 Manning and Pischke's falsification test

The NCDS provides test scores at ages 7, 11, and 16, obtained before primary school, before secondary school, and after secondary school respectively. In the true model age 16 test scores are a function of ability, type of secondary school (i.e. treatment of interest, here comprehensive attendance), and background variables:

$$Y_{16i} = \beta_0 + \beta_1 A_{16i} + \beta_2 Comp_i + \beta_3 B_i + \epsilon_{16i}. \tag{10}$$

Yet, all dimensions of ability are hardly observable in practice. In order to address the problem of missing confounding variables in the estimation of educational outcomes, most value-added specifications model outcomes as a function of prior student performance, school characteristics and other background covariates (Galindo-Rueda and Vignoles, 2005):

$$Y_{16i} = \alpha_0 + \alpha_1 Y_{11i} + \alpha_2 Comp_i + \alpha_3 B_i + \eta_{16i}. \tag{11}$$

A similar specification is assumed to hold for pre-secondary school educational outcomes at age 11. As a 'falsification test', Manning and Pischke control for comprehensive attendance, which should presumably not be a predictor of pre-secondary school outcomes:

$$Y_{11i} = \gamma_0 + \gamma_1 Y_{7i} + \gamma_2 Comp_i + \gamma_3 B_i + \epsilon_{11i}. \tag{12}$$

If the estimate for β_2 is significantly different from zero in the pre-treatment sample, then there might be misspecification issues in 12, and by similarity in 11 too. Manning and Pischke (2006) suggest that the estimate for α_2 is not picking up the treatment effect as intended, but that it suffers from selection bias, due to omitted confounders, and measurement error in test scores. The authors offer two alternative explanations of why $\beta_2 \neq 0$, which may rule out selection bias. The first one refers to 'coaching effects' experienced at age 11 by pupils in selective areas, implying that those attending a comprehensive school in the future experience a treatment effect of comprehensive before receiving treatment, since they do not receive training for the 11-plus, which pupils living in selective areas are likely to receive. The second one amounts to bias caused by measurement error in age 7 maths score.

Children in selective areas might receive 11-plus coaching both in primary schools, which arguably have an interest in having high 11-plus pass rates; and at home, if their parents are concerned enough about their children's education and able to afford it. Manning and Pischke hold that, while plausible, this explanation is unlikely to account for all of the bias. Results in fact do not change when the same models are estimated by the authors including only individuals from mixed LEAs, in which differences in 11-plus coaching are assumed to be smaller or minimal, since it is not clear a priori who might attempt the 11-plus.

Now consider the issue of measurement error in age 7 maths score, which might cause biased estimates of the coefficients of the sample equivalent of 12. The authors check for measurement error in maths scores at age 7 by comparing OLS coefficients of the regression presented to IV coefficients. The latter are obtained by instrumenting age 7 maths score on age 11 scores for other tests (reading, reasoning etc). Their findings do support the presence of measurement error, although note that the IV strategy might violate exclusions restriction, since age 11 scores will be related to age 11 maths score.