ESSAYS ON THE ECONOMICS OF HUMAN CAPITAL AND HEALTH

Chiara Pastore

PhD

University of York
Economics
December 2019

Abstract

This thesis consists of three essays on the topic of early determinants of human capital and health.

Chapter 2 explores the average effect of selective schooling, assigning pupils to different secondary schools by ability, on adult health, well-being and labour market outcomes, adding timely evidence to the debate on the reintroduction of selective schools in England. Selection bias is addressed by balancing pre-treatment characteristics, followed by OLS and IV regressions. Findings show that selective schooling does not affect most measures of long-term health and well-being. However, it affects educational aspirations and it raises adult wages for both high- and low-ability pupils, compared to a mixed-ability system.

Chapter 3 analyses the effect of selective schooling on long-term human capital for the marginal admitted pupil, comparing attendance to an academic school to its vocational alternative. Identification relies on a fuzzy regression discontinuity design, using proxies of entry test scores for selective secondary schools in England. Discontinuities in school assignment are estimated directly from the data. For the marginal admitted student, selective school attendance positively affects educational attainment, but this effect is conditional on having a favourable background. Other adult labour market and health outcomes are not affected.

Chapter 4 investigates the effect of birth order of children in the family on risky behaviours and non-cognitive skills in adolescence. The paper uses a mother fixed-effect strategy to account for the endogeneity of fertility decisions and data from a UK household panel. Having older siblings is linked to a higher probability of engaging in early drinking, drug use and skipping school, and to lower non-cognitive skills. The link is stronger for boys and higher socio-economic status families. Differences in parental investments and the influence of older siblings explain part of the observed birth order effects.

To Grazia, Francesco and Lorenzo.

Contents

List of Tables

List of Figures

Acknowledgements

A number of people have helped me along the tortuous PhD journey, and I am grateful to each one of them.

I owe my first thank you to my supervisors, Professor Andrew M. Jones and Professor Nigel Rice, for sharing their invaluable knowledge and experience, discussing ideas and patiently providing feedback at every step of the way. They have encouraged an open and stimulating environment to discuss research with the other HEDG students and I feel lucky for having been part of this group. I also wish to thank Professor Thomas Cornelissen for his acute observations and valuable comments throughout the last three years. I am deeply grateful to Dr Stefanie Schurer and Dr Agnieszka Tymula, for their mentorship and encouragement during my research stays in Sydney and beyond. I also want to thank the whole team at the Health Economics Research Unit in Aberdeen, especially Dr Verity Watson, Dr Patricia Norwood and Professor Paul McNamee, for giving me my first opportunity to have a go at research work. A big thank you also to Annette Johnson, Dr Michael Shallcross and Vin McDermott for their constant support.

I dedicate this thesis to my mother Grazia, for being on my team no matter what and for being my pillar throughout life. By choosing clever toys early on, she invested in my human capital before any of the research on early life circumstances told her to. To my father Francesco, for reminding me to take a deep breath and do my best to follow through with each project until the end. And to my brother Lorenzo, for always saving me from taking myself too seriously. I also want to thank the rest of my family for their continuous love and support, and make a special mention to my grandmothers Tonia and Anna. We've come a long way since Anna's school book was eaten by a goat while she was trying to do her homework.

I would not have got through the last few years as happily as I have, had it not been for the great friends I found along the way. Thanks to Junius, Sabrina, Anamaría, Fabio, Kezia and Maria Lucia, my life is richer because of them. I also wish to thank my colleagues, and especially Iván, Sara, Laia, Gloria and Vigyan for the numerous conversations and their supportive presence around the department. I also thank Andrea and Paulina for the incredible conversations and dinners wherever we are, I hope there are many more to come.

A special thank you goes to my second family, who made my daily routine more eventful and thanks to whom York became a true home: Laura, Ángela, Álvaro, Mahy, Álvaro again and Felipe, thanks for keeping me sane. I also need to thank Laura for sharing this journey, the morning coffees and the positive energy boosts, Álvaro for cheering up the greyer days and Ángela for sweetening the bitter ones. Finally, I wish to thank Valeria for her constant encouragement, as well as sharing adventures and life dilemmas while we grow up together; my long-time friends Gilda and Francesca, whom I can always count on; my university friends Valeria, Rossy, Nat and Yvonni; and Heleen, with whom the passion for health economics started. My final thank you goes to Alberto, for having lovingly stood by my side cheering me on, and for being a constant source of inspiration.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. Where individual chapters were co-authored with other researchers, this is indicated with the necessary specifications in this declaration. The work was financially supported by the Economic and Social Research Council (grant number ES/J500215/1). Finally, I declare that the funders, data creators and UK Data Service have no responsibility for the contents of this thesis.

Chapter 2 is co-authored with Professor Andrew M. Jones and Professor Nigel Rice, and an earlier version of it is available as working paper number 18/32 in the Health, Econometrics and Data Group (HEDG) Working Paper Series, under the title *"Tracking pupils into adulthood: selective schools and long-term well-being in the 1958 British cohort"*. I am the lead author of the paper, having developed the initial idea, designed and conducted all the empirical analysis, written the paper and done the necessary revisions. My co-authors contributed at various points with comments on design and methodology. I have presented earlier versions of the paper at the University of York, the NCDS 60 Years of Our Lives conference, University College London, the 2018 Royal Economic Society Junior Symposium, University of Sussex, the 4th Workshop on Applied Health Economics and Policy Evaluation Workshop, IRDES-Dauphine and the 3rd IZA Workshop: The Economics of Education, IZA.

Chapter 3 is co-authored with Professor Andrew M. Jones, and an earlier version of it is available as working paper number 19/08 in the HEDG Working Paper Series, titled *"Human capital consequences of missing out on a grammar school education"*. I am the lead author, having contributed to the design and implementation of the analysis, written and disseminated the paper. My co-author had the initial research idea and contributed to the design of the empirical strategy. I have presented an earlier version of this paper at the 2019 Royal Economic Society Conference, University of Warwick, the Health Economics Research Unit, University of Aberdeen, the 2019 Essen Health Conference, University of Duisburg-Essen, the XXVIII Meeting of the Economics of Education Association, University of

Las Palmas, and the 6th EuHEA PhD Student-Supervisor and Early Career Researcher Conference, Católica Porto Business School. The paper was also covered by the national press in newspaper articles by The Times (13/04/2019) and online by the Mail Online (13/04/2019). Further, it was featured in the Royal Economic Society media briefings and in a master's lecture series at the University of Bath.

I am the sole author of Chapter 4. I have presented an earlier version of this paper as a poster at the 4th Workshop of Labour and Family Economics, University of York.

Chiara Pastore
York, December 2019

Chapter 1

Introduction

This thesis comprises three papers on early determinants of human capital, with a special focus on health and well-being. Human capital is a notion that finds its popularity in economics starting from the seminal works by Mincer (1958) and Becker (1962). Economics used to be the science of monetary wealth and growth but, since the 1960s, a large branch of economic research has turned to education, skills and health as measurable objectives of what is taken to be a good life. These dimensions constitute human capital, which these first works modelled as a stock individuals can augment or deplete with their investment decisions. The determinants of such decisions, and the environmental constraints they are made in, have raised and continue to raise considerable interest, since simple observation of the world around us highlights differences in human capital and quality of life that may be preventable. Given that equality of opportunity is an intrinsic aim of many societies around the world, research on the determinants of human capital has become paramount, as it gives governments the evidence and tools to provide a more equal starting ground for all individuals to flourish.

Recognising the importance of early life environment to understand individual decisions and related constraints, this thesis focuses on schooling and on the family as two determinants of human capital accumulation. Educational policy may be one of the most effective tools to improve the quality of individuals' opportunities, regardless of their background. A central issue in the provision of public education is whether and how to tailor the curriculum around pupils' ability. Thus, Chapters 2 and 3 explore the long-term human capital consequences of a school system that operates selection on the basis of ability in England. Selective schools have been on and off the UK policy agenda for decades since the 1940s. Yet, this work comes at a particularly policy-relevant time, since the UK government spent £50 million on a new fund for selective schools expansion in 2018, announcing the following £50 million round for 2019 as part of a bigger project worth £200 million in total (UK Department for Education, 2019a).

On the other hand, family background is often seen as a panacea for differences in lifetime health and human capital. However, disentangling the actual causal paths in this black box of family background effects can be challenging, as there are multiple aspects to consider, including but not limited to parents' education and socio-economic status, parenting styles and investments, genetics, health, sudden shocks, and other cultural and intangible factors. Chapter 4 explores the role of birth order of children in the family to explain adolescent risky behaviours and non-cognitive skills. It additionally investigates whether birth order effects can be linked to parental investments and interactions among siblings. Assessing the role of specific educational policies and family circumstances 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. 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. The following paragraphs outline the motivation and contribution of each chapter.

Chapter 2 examines the long-term human capital consequences of the 1960s transition from a selective to a mixed-ability system of secondary schooling in England and Wales. The aim is to observe the average effect of this change and understand its impact on human capital, in light of the current government policy, which is increasing the number of publicly funded selective school places. The rationale for using an older dataset to answer a new policy question comes from the unique opportunity to observe long-term consequences of a given measure. In this particular context, a second motivation is that selective schooling was still largely present at the time, giving a reasonable sample size of pupils affected by selection, before mixed-ability schools became the norm (Bolton, 2017). The literature on the effect of selective schools follows two main strands. Some studies compare outcomes for selective versus non-selective systems (Basu et al., 2018; Burgess et al., 2017, 2019; Galindo-Rueda and Vignoles, 2005; Jones et al., 2012), while others compare outcomes for individuals who pass selection, versus individuals who do not, within the same selective system (Abdulkadiroglu et al., 2014; Burgess et al., 2017; Clark, 2010; Del Bono and Clark, 2016). Chapter 2 falls under the former category, comparing average outcomes under a system where pupils are assigned to selective 'grammar' schools upon passing an ability test, or to vocational 'secondary modern' schools otherwise, to those in a non-selective system, where all pupils go to mixed-ability 'comprehensive' schools.

The study uses data from the National Child Development Study, an ongoing longitudinal study following a cohort of individuals up to age 60. This chapter contributes to the field by using a broad range of human capital outcomes. These add new dimensions to existing knowledge, but at the same time aim to explain results from previous literature, mostly concerned with educational and labour market outcomes. The chapter analyses school and work aspirations at the end of compulsory schooling age, which can work as channels for later educational achievement and earnings. It looks at earnings and employment, but also at non-monetary measures of adult well-being, such as adult life and job satisfaction, self-efficacy and crime participation. Finally, it evaluates long-term effects of school type on adult self-assessed health and biomarkers for risk of cardiovascular disease up to age 55, based on the idea that more and better schooling improves an individual's health production function (Grossman, 1972). A move to a more selective system today is expected to increase average peer ability and school quality for the additional pupils admitted to grammar, while these would decrease for lower-ability pupils who do not pass selection. To account for this key difference, the second contribution of Chapter 2 is to estimate two separate average effects of selective schooling, splitting the sample of pupils along the cognitive ability dimension. Thus, grammar pupils are compared to comprehensive pupils who, given their ability scores, would have gone to grammar, had they experienced the selective system. Similarly, secondary modern pupils are compared to comprehensive pupils who are closest to them in terms of ability. In practice, similarity is achieved through entropy balancing (Hainmueller, 2012), matching individuals on their pre-treatment characteristics. The strategy relies on the validity of the conditional independence assumption, the implications of which are discussed, also in light of previous criticisms to the reference literature (Manning and Pischke, 2006). Results highlight that grammar schools positively affect academic aspirations and employment prospects for their pupils, but they also lower adult life satisfaction, compared to their mixed-ability counterparts. Secondary modern pupils, of lower cognitive ability, display higher average adult wages and, in some specifications, higher self-efficacy than their comprehensive counterparts. However, regardless of the cognitive ability level, attendance to selective schooling is not significantly linked to other measures of health and human capital later in life.

Chapter 3 explores the effect of attending a grammar school in the 1970s, compared to not being admitted, on long-term human capital outcomes, within the selective system. Like the second category of studies mentioned above, this chapter focuses on the marginal additional student admitted, estimating the average effect of going to grammar school for a restricted group of individuals who score near the pass mark, with secondary modern as the alternative. Focusing on this group has two advantages. First, these individualswould have been the ones most likely to be affected by an expansion in available grammar places. Second, since individuals in this region have similar cognitive ability, it allows estimation of the effect of schooling while keeping ability constant. Accounting for the results of Chapter 2, this can help further reduce differences in outcomes due to background ability, and isolate the effect of grammar, compared to secondary modern. Again, the data are from NCDS and outcomes span the education, labour market and long-term health spheres, adding to existing knowledge on the effect of selective schools for the marginal student, using for the first time a sample of pupils from several regions in England to answer this specific question.

Identification of the local average treatment effect for individuals who score near the pass mark is achieved via a fuzzy regression discontinuity design, assuming that the probability of attending grammar varies discontinuously at the pass mark, and that all other individual characteristics related to the outcomes are smooth functions of the ability score. An innovative element of the study design, compared to previous literature, is that information on the pass mark, set at local education authority level, is inferred from the data, by looking for structural breaks in the probability of grammar school attendance as a function of the ability score (Bai, 1997; Porter and Yu, 2015). Originating from the financial time-series literature, this method to locate discontinuities has been used in the labour economics literature only a handful of times. By applying this method to a context where limited data has always represented an issue for identification of a causal effect, Chapter 3 can be an example of how to proceed in policy evaluations in situations where, in spite of the limited data, it is known that a discontinuity in treatment assignment exists. Several robustness checks are provided to ensure that the discontinuity is not spurious. Results show that in a selective system, attendance to the higher quality school only matters for educational achievement, although this is conditional on having a favourable family background. These findings are discussed in light of other literature, recognising that any significant effects of type of school could be concentrated either at the top or at the bottom of the ability distribution. However, an expansion in grammar school places at the time would have led to higher educational attainment for advantaged groups, and no other advantage in terms of human capital for the additional admitted pupils.

Chapter 4 shifts the focus to family background, analysing the effect of children's order of birth in the family on the probability of engaging in risky behaviours and on non-cognitive skills in adolescence. Several theories from psychology postulate the role of birth order as a determinant of individual behaviour, with personality taking shape as the child reacts to the surrounding environment (Adler, 1928; Rohrer et al., 2015; Sulloway, 2001). Additionally, birth order is an attractive feature to explore, as research on human development looks increasingly earlier for the roots of human capital, highlighting the importance of the time around birth and even earlier *in utero* for long-term individual outcomes (Akresh et al., 2014; Almond and Currie, 2011; Persson and Rossin-Slater, 2018; Schwandt, 2018; Von Hinke Kessler Scholder et al., 2014). In the economic literature, most papers on birth order have focused on the link with childhood cognitive skills and educational achievement, while a few studies have analysed long-term consequences for adult health, occupation and personality traits (Black et al., 2016; Black et al., 2018; Lehmann et al., 2018). Adolescence remains largely unexplored, although it is universally recognised as a crucial time for skill and healthy habit formation, as well as the first time individuals make decisions by themselves, instead of relying entirely on their parents. As a secondary aim, the chapter then explores parental behaviour and imitation patterns among siblings as explanations for behavioural differences by birth order. Unpacking the relationship between birth order and adolescent behaviour can thus be a starting point to devise strategies aimed at healthier decisions at a key time for the individual's physical and mental development.

The main challenge in the literature on birth order effects consists of separating the effect of birth order from that of other family characteristics, including family size. Chapter 4 uses data from a panel of UK households exploiting information on several siblings to net out the portion of the variation in behaviour due to specific family traits, and isolate, to the extent that it is possible, the effect of birth order. The evidence presented points towards a higher probability of engaging in risky behaviours and to lower non-cognitive skills in adolescence for children with older siblings. Some heterogeneity is found, indicating larger birth order effects for boys compared to girls, and for families of higher socio-economic status. It is further shown that parental interest in school and support with homework decline with birth order. These variables explain a significant portion of the decrease in non-cognitive skills and in some of the risky behaviours considered. Sibling interactions are also shown to be important to understand birth order effects.

Finally, Chapter 5 concludes by discussing the contribution of this work in light of its results, including the significance of the findings for policy, and by touching on open questions and future avenues for research.

Chapter 2

Tracking pupils into adulthood: selective schools and long-term human capital

Tracking pupils into adulthood: selective schools and long-term human capital

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

Abstract

We explore the effect of selective schooling, where students are assigned to different schools by ability, on aspirations, well-being, labour market outcomes and adult health. In the 1960s, England and Wales experienced a transition from a selective to a non-selective secondary schooling system. We distinguish between two effects of the introduction of mixed-ability schools, recognising that average school quality and peer ability decreased for high-ability pupils, and that they increased for low-ability ones, following the transition. We address selection bias by balancing individual pre-treatment characteristics via entropy balancing, followed by OLS and IV regressions. Selective schooling marginally raises hourly wages, compared to a mixed-ability system, while it affects school aspirations and life satisfaction differently, depending on the ability level of the school. However, most measures of long-term health and well-being are not affected.

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

2.1 Introduction

Tracking students by ability into different schools at a young age is a controversial policy. On the one hand, it can be seen as a way to improve learning and teaching efficiency, by catering for different abilities separately. The practice is also argued to reduce socio-economic inequalities, since selection into prestigious institutions is based on academic talent, regardless of family circumstances in principle. On the other hand, such systems have been shown to favour children from affluent backgrounds, who are generally more supported by their families and more prepared to take entry tests (Burgess et al., 2018; Cribb et al., 2013). If less advantaged students are more likely to be excluded from the upper tracks, then selective schooling could have detrimental effects on the pre-existing inequality gap, going against the equality of opportunity principle advocated by many modern societies (Oakes, 2005; OECD, 2016)[1].

Several countries incorporate selection by ability in their secondary schooling systems, including Australia, England, France, Germany, the Netherlands, Switzerland and the United States. Research on the effects of tracking on educational and labour market outcomes is abundant. Attending an upper track school is generally linked to better educational outcomes, but the presence of tracking is also associated with higher inequality in education and earnings, often leading to low social mobility (Brunello and Checchi, 2007; Burgess et al., 2017, 2019; Hanushek and Wößmann, 2006). Other non-monetary benefits of education, such as health and well-being, have received less attention in this literature. In this paper, we assess the human capital effects of selective versus non-selective schooling, by looking at long-term health, well-being and labour market outcomes for a British cohort.

Our paper is timely in providing evidence on the long-term consequences of selective schooling, given its recent expansion in England. In 2018, the UK government allocated a first £50 million investment towards new selective school places, and a second round is under way in 2019, at the time of writing (UK Department for Education, 2019b). In order to learn about the long-term consequences of selective schooling, we exploit the comprehensive schooling reform implemented in England and Wales in the 1960s, which caused some areas to transition from a selective to a non-selective system of secondary education earlier than others. The empirical analysis relies on data from the National Child Development Study (NCDS), a British cohort study of individuals born in March 1958, allowing us to follow their lives to date. Depending on the area they lived in

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

at the time, NCDS children were exposed to either a selective or a non-selective system. In selective areas, an entry test determined whether a pupil was offered a place in a selective grammar school, representing the more academic track, or in a vocational secondary modern school, the main alternative for low-scoring pupils. In the non-selective system, schools were converted into or created as comprehensive, institutions receiving pupils of all abilities. Attendance at different school types exposed pupils to different curricula, teacher quality and peer ability, thus offering an opportunity to explore the long-term human capital effects of variation in school quality.

The transition to comprehensive schooling presumably affected pupils differently, depending on their cognitive ability. Tracking affects the quality of the school but also the average peer ability each pupil is exposed to. Thus, for high-ability pupils, the transition to the non-selective system lowered school quality and average peer ability, while for low-ability pupils these increased. Our empirical approach attempts to mirror the differential effects of the transition by estimating two treatment effects. On the one hand, we explore long-term effects of attending grammar, compared to comprehensive, for pupils of high cognitive ability. On the other, we investigate the effect of attending secondary modern, compared to comprehensive, for pupils of lower ability. The other advantage of separating treatment effects is that we make treatment and control groups more comparable. To this end, we additionally implement entropy balancing to close the gap in pre-treatment characteristics, thus increasing confidence that we are addressing the issue of endogeneity of school quality. The balancing exercise is followed by parametric regressions for a rich set of outcomes. Throughout the paper, we refer to our estimates as 'treatment effects', since our aim is to construct a quasi-experimental setting to evaluate selective schooling. However, recognising that we cannot definitively rule out bias due to unobservables, we clearly state the relevant assumptions and discuss the implications of their failure, the main consequence being that our estimates would yield a correlation of type of school with the outcomes, rather than a causal effect.

We build on the literature exploring health impacts of the comprehensive reform in the 1960s (Basu et al., 2018; Jones et al., 2011, 2012). However, our study is the first to consider two separate treatment effects by splitting the sample by cognitive ability, in order to assess the link of selective schooling to biometric markers for cardiovascular disease and stress, as well as to several dimensions of well-being and human capital in adulthood. We find that type of secondary school attended does not affect most of our adult health and well-being outcomes, with some exceptions. Attitude towards school is positively linked to grammar school attendance and negatively to secondary modern attendance, compared to comprehensive school attendance and for similar ability levels. This could be a channel for educational outcomes, found to be significantly linked to type of school by Burgess et al. (2017) and Guyon et al. (2012). The second exception are labour market outcomes, which are better for grammar pupils, confirming previous studies on earnings (Burgess et al., 2019; Del Bono and Clark, 2016). Somewhat surprisingly, we find marginally better wages also for pupils in secondary modern compared to equivalent comprehensive pupils, and offer some hypotheses as to why this may be the case. We additionally note lower adult life satisfaction for grammar pupils and higher self-efficacy for secondary modern pupils, compared to their mixed-ability counterparts. Our findings are based on historical data and concern the consequences of the grammar school system for today's generation of 60-year-olds. Assessing the extent to which they may apply to the reintroduction of selective schools today would require a consideration of how the education and labour market systems have evolved over time, but this is outside of the scope of this work.

The paper develops as follows. Section 2.2 reviews the history of selective schooling in England and Wales and the existing relevant literature. Section 2.3 describes the data and Section 2.4 outlines our two-step approach to estimate the relationships of interest, combining entropy balancing with parametric regressions. Section 2.5 presents the main results, along with appropriate robustness checks. Section 2.6 discusses these findings and Section 2.7 concludes.

2.2 Background

2.2.1 Selection by ability in England and Wales

The origins of tracking in the British school system go back to the 1944 Education Act, which established the reorganisation of state secondary schools by local education authorities (LEAs) in a tripartite system, comprising grammar, secondary modern and technical schools. 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, around age 11. The 11-plus was set at LEA level, so difficulty and entry score varied across the country. Grammar schools admitted on average pupils scoring in the top 25% of the cognitive ability distribution in their local area (Bolton, 2017). Entry tests consisted of different modules, including mathematics, English and verbal and non-verbal reasoning. Panels of teachers and LEA representatives allocated grammar places according to test scores, capacity constraints, proximity and other considerations. Although parents had the opportunity to appeal if they disagreed with the outcome, for the

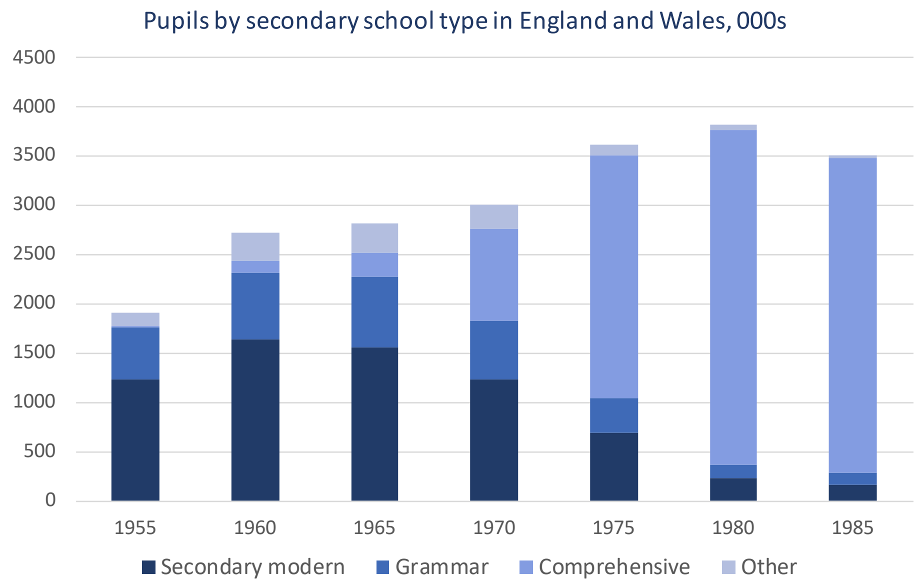


Figure 2.1: Number of pupils by school type over time. Source: Bolton (2012). Technical schools included in 'Other'.

great majority of pupils not passing the exam meant they would attend secondary modern schools, less academically demanding, geared towards trades. The third type, technical schools, were for vocational training, did not require an exam and were not prevalent.

Given the growing dissatisfaction with the allocation system in state schools, with Circular 10/65 in 1965, the Labour government started promoting a phase-out of the selective schooling system (Kerckhoff, 1996). 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 (see Figure 2.1) and generally slower in areas with a Conservative political majority. In 1998, with the School Standards and Framework Act, the Labour government outlawed establishment of any new schools that selected pupils by ability. At the time of writing, 163 grammar schools exist in England, attended by approximately 167,000 pupils, while Welsh schools are wholly comprehensive (Bolton, 2017). The first round of the Selective Schools Expansion Fund, launched in 2018, funded expansion projects in 16 existing grammar schools (UK Department for Education, 2019b), thus begging the question of the impact of a return to a more selective education system for present and future generations of pupils.

2.2.2 Related literature

The 1960s comprehensive reform in England and Wales offers an opportunity to evaluate long-term effects of the transition from a selective to a non-selective secondary schooling system. Yet, the lack of a clear roll-out of the reform has made it difficult to isolate the effect of school type on individual outcomes from other confounding factors. Since type of school and adult outcomes could both be influenced by characteristics that are unobservable or difficult to measure, such as individual ability or parental investments, the standard problem of endogeneity in estimating returns to education also applies to this context (Angrist and Krueger, 1991). The literature has dealt with this issue in different ways, mainly to estimate effects on earnings and educational achievement. Using NCDS data, Galindo-Rueda and Vignoles (2005) estimate the effect of comprehensive attendance on test scores at age 16, by controlling for prior test scores, in a so-called value-added approach (Todd and Wolpin, 2003). In a second instrumental variable (IV) specification, they 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 comprehensive reform reduced educational achievement for more able children only. The validity of this type of analysis was put under scrutiny by Manning and Pischke (2006), who criticise value-added approaches, comparing outcomes for pupils in selective and comprehensive areas (also found in Harmon and Walker (2000), Jesson (2000), and Kerckhoff (1986), among others). They argue that adding pre-secondary school outcomes as controls is not sufficient to remove endogeneity, since the two groups are too fundamentally different. This is demonstrated by showing that a spurious treatment effect is found when running a placebo regression of pre-secondary school test scores on an indicator for comprehensive school attendance. The IV strategy proposed by Galindo-Rueda and Vignoles (2005) does not solve this problem either. This is taken as evidence against their results, a conclusion 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 in the NCDS cohort disappears.

A number of studies have used alternative methods that are more robust to the criticisms advanced above. Maurin and McNally (2009) use two cohorts of individuals born twelve years apart to compare the effects of selective and non-selective systems of education in England. They find that, while attending grammar school is linked to better individual outcomes, the 1960s transition to non-selective schooling led to an increase in average educational outcomes, with larger benefits observed for lower socio-economic status individuals. In a different study, Burgess et al. (2017) analyse the effects of selection both within selective areas and across areas with different levels of selectivity, using administrative data from the National Pupil Database. Within selective areas, grammar attendance increases pupils' chances of accessing and completing higher education. Secondly, by matching selective and non-selective areas to ensure similarity in area characteristics, they find that high ability children who do not get into top-tier schools do worse in selective areas. Moreover, since they show that access to grammar is strongly linked to higher socio-economic family background, they conclude that the grammar school system does not promote social mobility. With similar matching methods at the area level, Burgess et al. (2019) then investigate the impact of selective schooling on the earnings distribution, finding that inequality in average hourly wage is significantly higher in selective areas.

Another stream of the literature has used regression discontinuity methods, estimating the effect of the upper academic track, based on pupils scoring close to the entry cut-off. Using data for the East Riding of Yorkshire, a region in the UK, Clark (2010) finds only a small positive effect of attending grammar school on test scores, while a slightly larger and positive effect is observed for university enrolment. In another study, Del Bono and Clark (2016) estimate the impact of Scottish elite schools on educational attainment, income and fertility for the marginal student[2]. Elite schools increase several measures of educational attainment, while small effects on labour market outcomes (positive) and fertility (negative) are found in women, but not in men. With similar methods, Guyon et al. (2012) evaluate the effects of an expansion of grammar school places in Northern Ireland, and find that it increased average educational outcomes when looking at the whole distribution. However, the expansion also decreased average outcomes for non-grammar school pupils.

Health effects of the comprehensive reform are somewhat less explored in the literature. Using NCDS data, Jones et al. (2012) show that the distribution of health outcomes for grammar pupils strictly dominates that for comprehensive and secondary modern pupils. However, the study also finds that the association of school type with self-reported health outcomes is mostly insignificant when accounting for pre-school characteristics via probit and linear models. In another relevant paper, Jones et al. (2011), evaluate the impact of educational attainment and school attributes on self-reported health behaviours and outcomes in NCDS. Although they do not directly look at the impact of type of school on outcomes, they find a stronger association of educational attainment with health behaviours

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

for pupils in low-ability schools, and with mental health in high-ability schools. Finally, in a more recent study using the NCDS, Basu et al. (2018) explore the effect of selective versus comprehensive schooling on three dimensions of adult health and on smoking. They focus on understanding heterogeneity by ability, by estimating marginal treatment effects along the cognitive ability distribution. Using percentage of comprehensive schools in the individual's LEA in 1969 as a continuous instrument, they find a negative effect of the move to comprehensive schools on depression only, magnified for individuals with lower non-cognitive skills. Furthermore, they rule out 'essential heterogeneity' in the effect due to unobserved factors, since their effect estimates achieved with a local instrumental variable method are similar to those obtained via standard OLS and IV methodologies[3]. Overall, this literature suggests that accounting for differences in prior ability is key to estimating an unbiased effect of selective schooling.

Our analysis builds on this previous work and makes two important contributions. First, we split treatment effect into two along the cognitive ability dimension. We do this to acknowledge that the transition from a selective to a comprehensive system represented a different treatment for individuals of high and low cognitive ability. The entropy balancing algorithm increases comparability between treated and control groups prior to treatment, 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 life. The rationale for this broad scope is to better understand the paths leading from education to adult inequalities in health, income and general well-being, with the inclusion of biometric outcomes for the first time in this literature.

2.3 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, 98% of all individuals born in that week. Approximately 9,000 were retained at the most recent wave in 2013 (Brown et al., 2016). Following the birth survey, 9 further sweeps have been undertaken to date, at ages 7, 11, 16, 23, 33, 42, 46, 50 and 55, plus the collection of biomedical

[3]The local instrumental variable (LIV) method used by Basu et al. (2018) is meant to account for unobserved heterogeneity by estimating marginal treatment effect, relying on the assumption that the distribution of both observable and unobservable characteristics are the same for very small variations in a given continuous instrumental variable. More on their methods in Basu (2014) and Heckman et al. (1999).

samples and data at age 45[4]. Due to differences in the schooling systems between countries, we only include individuals going to school in England and Wales in the analysis[5]. The key variables for the present study are described below.

2.3.1 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 when estimating the effect of school quality (Manning and Pischke, 2006). In addition to individual characteristics, family background covariates include mother's interest in child's education (expressed on a 0-4 scale), father's employment status and socio-economic 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 et al., 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 11 also includes whether the child goes to an independent primary school; child's happiness at school reported by parents; whether the child will go to school or study after minimum school-leaving age. Finally, local area characteristics, based on LEA of school attended in 1974, were retrieved from the 1971 Census (full list in Table 2.2).

2.3.2 School

The 1958 cohort started secondary school in 1969, during the transition to the comprehensive system, meaning that cohort members experienced one of two different secondary school systems, selective and non-selective. Figure 2.2 reports the percentage of comprehensive pupils by LEA at the time, showing considerable variation across the country. 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 independent schools; technical (0.5%), and others (2.2%) (including all age, educationally subnormal, and other special needs). For our analysis, we consider only the first three categories, individu-

[4]A detailed breakdown of the data collected for each sweep can be found in the cohort profile by Power and Elliott (2006), and online at https://cls.ucl.ac.uk.

[5]Welsh individuals overall represent under 5% of our sample.

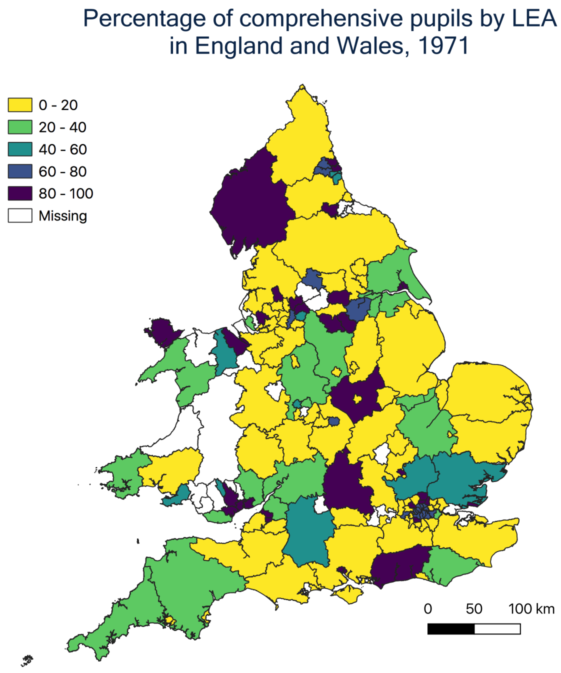


Figure 2.2: LEAs with higher percentages of comprehensive pupils are more ahead in the transition to the mixed-ability system and viceversa. Sources: Byrne and Williamson (1976) and Comprehensive School Committee (1971).

als going to state schools for whom we have all covariates of interest, leaving a sample of 7,694 individuals: 1,040 grammar, 1,991 secondary modern and 4,663 comprehensive pupils. The data on LEA of the school was obtained under special licence. LEA percentage of comprehensive pupils aged 13 in 1971 (corresponding to the NCDS cohort) was retrieved from the 1971 edition of the Comprehensive School Committee (CSC) Journal[6].

2.3.3 Ability

Cognitive skills were assessed through numeracy, reading, verbal and non-verbal tests at ages 7, 11 and 16. Thus, tests were administered during primary, just before secondary and just after secondary school respectively (see Figure 2.3). Following existing literature, we group test scores to obtain three indicators for cognitive ability, one for each age, by implementing principal component analysis (PCA) (Cawley et al., 1997; Galindo-Rueda and Vignoles, 2005; Jones et al., 2011). For simplicity of interpretation, we then convert the three PCA indices to variables bounded between 0 and 1. PCA captures the variation in the data, while avoiding multicollinearity issues that would arise if all the test scores were

[6]Most of these figures 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).

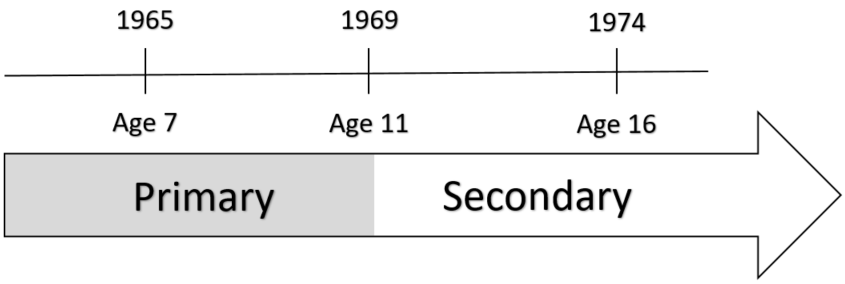


Figure 2.3: The graph shows the timing of each cognitive ability test undertaken by NCDS cohort members in relation to their attendance to primary and secondary school.

included as regressors in the model. More details on the construction of the three indices can be found in Appendix Section A.3. As noted by Basu et al. (2018) and Jones et al. (2011), age 11 tests closely resemble the three components of the 11-plus: mathematics, reading, verbal and non-verbal reasoning. When 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. An index based on these factor loadings is therefore going to mirror the 11-plus, where equal weights are given to its different components. We additionally construct a cognitive rank variable, ranking NCDS individuals by their measured cognitive ability at age 11. This is calculated separately for children attending the selective system (grammar and secondary modern schools) and the mixed-ability system (comprehensive schools). Finally, pre-secondary school non-cognitive skills are proxied by the Bristol Social Adjustment Guide (BSAG) score, grouping teachers' answers on twelve dimensions of child behaviour at school, measured at age 11. The twelve attributes are measures of social maladjustment and include unforthcomingness, withdrawal, depression, anxiety for acceptance by adults, hostility towards adults, 'writing off' of adults and adult standards, anxiety for acceptance by children, hostility towards children, restlessness, 'inconsequential' behaviour, miscellaneous symptoms and miscellaneous nervous symptoms[7]. For the present analysis, the BSAG score is converted to a variable bounded 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, indicating no behavioural problems.

[7]More details on the measure and questionnaire used can be found in Shepherd (2013).

2.3.4 Outcomes

Well-being and labour market measures

In order to assess short-term impact of secondary schooling, we look at aspirations related to school and work measured at age 16, just after secondary school, as potential determinants of future achievements. School aspirations is a dummy variable equal to 1 if the individual intends to stay at school beyond 16, the minimum school-leaving age. Work aspirations is also a dummy variable, indicating whether the individual aspires to personal and intellectual growth through a job. Adult well-being outcomes include life satisfaction, self-efficacy and positive feelings about one's job. These are based on the age 33 survey, and are all constructed via PCA, grouping answers to several questions[8]. 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 drug use dummy takes value 1 if the individual has ever tried any illegal drug. We also examine two labour market outcomes, each measured twice, at ages 33 and 50. The first is individual gross hourly wage, imputed from weekly, monthly or bi-monthly usual gross pay 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.

Survey health measures

The long-term impact of selective schooling is also assessed on a broad range of health dimensions in adulthood. Self-assessed health (SAH) is measured on a 5-point scale: Excellent (1); Good (2); Fair (3); Poor (4); Very poor (5). This measure has been shown to predict ill-health and mortality reasonably well and it has been validated across a variety of cultural contexts (Eriksson et al., 2001; Kaplan and Camacho, 1983). The 9-item Malaise Inventory, developed by Rutter et al. (1970), offers a measure of ill-health and discomfort, both physical and mental (Rodgers et al., 1999). The list of questions can be found in Appendix Section A.4. For ease of interpretation, in regressions we use binary variables equal to 1 for excellent or good SAH and for low malaise (defined as scoring lower than 2 on the malaise instrument). Both of these are measured at age 50. Mental ill-health is further measured at age 45 by the revised version of the Clinical Interview Schedule (CIS-R), developed by Lewis et al. (1992). It is expressed as a summary score ranging from 0 to 30 based on ten different areas: anxiety, appetite, con-

[8] For details on the questions used for variables constructed by PCA, see Appendix Section A.4.

centration/forgetfulness, depression, depressive ideas, fatigue, irritability, panic, phobias and sleep.

Biometric health measures

A body mass index (BMI) measure was constructed as weight in kilograms, divided by squared height in metres, using weight and height measured by a nurse at age 45. A healthy adult BMI ranges from 18.5 to 25kg/m^2. Individuals with BMI$<$ 18.5 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 Organization, 2017).

Blood samples taken at age 45 were used to measure lipids, clotting factors and inflammatory markers, obtained via special license access. Our outcomes include C-Reactive protein (CRP) (mg/L), fibrinogen (g/L) and triglyceride (mmo/L) levels, as well as cholesterol ratio (mmo/L), constructed as total cholesterol divided by high-density lipoprotein (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 assess the effect of education on an objective measure of the risk of presenting health problems in the future.

2.3.5 Attrition

As in most longitudinal studies, a concern when analysing NCDS data is that attrition can be non-random. If the probability of dropping out of the sample is related to variables correlated with the treatment or outcome, then estimates of treatment effect could be biased (Hausman and Wise, 1979). For each survey wave used, we therefore examine differences in average characteristics between our sample and dropped out individuals. We find that there are small but noticeable differences in average birth and childhood characteristics between these groups (see Table A1 in the Appendix). Individuals who dropped out 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. However, treatment status, school type, is observed at 16. The fact that all samples from age 16 onwards present hardly any differences in the average characteristics shown increases confidence that sample composition does not vary systematically in relation to key characteristics after this point in time. This is particularly reassuring for our analysis and in agreement with other literature (Case et al., 2005; Jones et al., 2011). Another important feature is that the percentage of pupils attending each type of school does not vary over time. Moreover, as noted by Dearden et al. (2002), even if lower ability and lower SES pupils were under-represented in the sample, controlling for such characteristics in our analysis reduces the potential for bias in treatment effect estimates.

2.4 Methods

2.4.1 Reference framework

Secondary school type, S_i is the key treatment of interest, and we assume it is a function of pupil's background, B_i, comprising family and individual characteristics, childhood abilities, A_i (this is particularly true in selective areas), and supply of places by type of school, SU_i.

$$S_i = S(B_i, A_i, SU_i) \tag{2.1}$$

The production functions for adult health and well-being outcomes, Y_i, depend on background, pre-secondary school ability, type of school, and local area characteristics[9].

$$Y_i = Y(B_i, A_i, S_i(.), LA_i) \tag{2.2}$$

In the framework, background B_i and ability A_i enter both the school-assignment function, Equation (2.1), and the outcome equation, Equation (2.2). If there are unobserved factors correlated with either background or ability, the standard OLS estimator of the effect of S_i in the empirical estimation of Equation (2.2) will be biased[10]. This issue represents the main challenge for identification of treatment effect in our context. In principle, establishing causal effects requires comparing treated individuals with credible counterfactuals (Heckman et al., 1997; Rubin, 1974). In this spirit, we split the sample into two, thus estimating two separate treatment effects.

Following the Neyman-Rubin framework, we denote two possible counterfactual outcomes for individual i as Y_i^0 in the absence of treatment, and Y_i^1 with treatment. On the one hand, we estimate the effect of going to grammar, compared to comprehensive, by comparing outcomes for grammar pupils to their

[9]We exclude from our framework any post-treatment variables, as these might bias treatment effect in the empirical estimation.

[10]To see why, suppose that the true relationship is $Y_i = \gamma_0 + \gamma_1 A_i^* + \gamma_2 S_i + \epsilon_i$, where true ability $A_i^* = A_i + U_i$, U_i is an unobserved term and $corr(S_i, U_i) \neq 0$. Since we can only observe A_i, we estimate $Y_i = \widetilde{\gamma}_0 + \widetilde{\gamma}_1 A_i + \widetilde{\gamma}_2 S_i + \epsilon'_i$, where $\epsilon'_i = \epsilon_i + \gamma_1 U_i$. Since $corr(S_i, U_i) \neq 0$, it follows that $corr(S_i, \epsilon') \neq 0$, and therefore the OLS estimate for $\widetilde{\gamma}_2$ will be biased (see Angrist and Pischke, 2009 for details).

counterfactual. These are comprehensive pupils who would have gone to grammar, had they gone through selection. The effect is an average treatment effect on the treated (ATT), conditional on control individuals providing a reliable counterfactual group:

$$ATT^G = E[Y_i^1 - Y_i^0 | G_i = 1]. \tag{2.3}$$

Similarly, we estimate the effect of going to secondary modern, compared to its counterfactual, comprehensive pupils who would have attended secondary modern, had they experienced the selective system:

$$ATT^{SM} = E[Y_i^1 - Y_i^0 | SM_i = 1]. \tag{2.4}$$

Since $E[Y_i^0|G_i = 1]$ and $E[Y_i^0|SM_i = 1]$ are never observed in practice, we need to build two counterfactual groups, one for each treatment. We do this 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 (2.2), and estimated using the weights obtained in the balancing procedure. Figure A1 in the Appendix illustrates how we construct our samples. The regressions rely on a set of assumptions, such as the functional form used and the specification of variables included in the model. While they 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[11]. Then, balancing covariates for treatment and control groups and using resulting weights in subsequent parametric regressions 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. The main remaining concern is related to unobservables, possibly confounding the relationship of interest. Estimation of an unbiased treatment effect in this context relies on the conditional independence assumption (CIA), expressed as $Y_i^j \perp S_i | \mathbf{X}_i$, with $j = 0, 1$. This assumption holds either if all characteristics correlated to treatment and outcome are observed and controlled for, or if by balancing on the observed characteristics, we also achieve balance on the unobserved characteristics. We test this

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

assumption in the placebo procedure illustrated in Section 2.5.6 and discuss the implications of relaxing it in Section 2.6.

2.4.2 Building a counterfactual: entropy balancing

Entropy balancing is implemented for the two separate 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. 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 the original sample[12]. Developed by Hainmueller (2012), the procedure assigns weights to the observations in the control group according to pre-specified conditions, in order to emulate the treatment group in terms of the moments and co-moments of specific covariates[13].

The covariates for the balancing procedure are selected based on their expected relationship to both treatment and outcomes (Caliendo and Kopeinig, 2008). The methodological literature highlights that this choice implies 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 variable related to outcome but not treatment will bias the estimate. In order to ensure that the variables are not influenced by treatment, which would also bias effect estimates, only pre-secondary schooling variables are used. We include cognitive test scores, BSAG scores as a proxy of non-cognitive skills, relative rank by cognitive score, mother's interest in child education and a dummy for high or middle-high father's SES. Except for cognitive test scores, measured at age 7, all variables are measured at age 11, just before starting secondary school. We prefer age 7 to age 11 cognitive ability scores, since the latter could be biased upwards in selective areas because of 'coaching effects' (Jones et al., 2011). This is the idea that students in selective LEAs score higher because they have been coached for this particular kind of test in view of the imminent 11-plus exam, meaning that age 11 scores do not reflect ability in the same way for pupils from selective and non-selective areas. Since the age 11 cognitive rank variable is constructed separately for selective and

[12] 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). These yielded matches of lower quality, and smaller sample sizes since observations outside common support are dropped. Nevertheless, final results adopting alternative matching procedures are not significantly dissimilar from the main ones presented.

[13] All empirical analysis is conducted using Stata 15. The Stata package *ebalance* allows for a straightforward implementation of the entropy balancing algorithm (Hainmueller and Xu, 2013).

non-selective pupils (see Section 2.3.3), the bias of coaching effects does not carry over to this variable. By balancing mean, variance and skewness of the five included covariates, as well as their pairwise interactions, we achieved close balance, without compromising the feasibility of the minimization procedure required for entropy balancing. As a sensitivity check, to address the concern that selective and non-selective areas are too different for comparison, we include fourteen local area characteristics from the 1971 Census 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[14].

2.4.3 Parametric regressions

We apply the weights obtained from entropy balancing to the control observations in parametric regressions. Assuming for each sample $j = GC, SMC$ a constant average treatment effect α^j, we estimate the following by ordinary least squares (OLS):

$$Y_i^j = \beta_0^j + \alpha^j S_i^j + \beta_1^j \mathbf{A}_i + \beta_2^j \mathbf{B}_i + \beta_3^j \mathbf{LA}_i + \epsilon_i^j, \tag{2.5}$$

with constant β_0^j 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 the vector of ability $\mathbf{A}_i$, including age 7 cognitive skills, age 11 non-cognitive skills and age 11 relative cognitive ability (the rank variable); the vector of individual background characteristics $\mathbf{B}_i$, including sex, ethnicity, family socio-economic status, childhood health endowment and primary school characteristics; and finally local authority characteristics $\mathbf{LA}_i$, while ϵ_i is a random error term. The whole set of pre-treatment covariates included is listed in Table 2.2. Among them, we pay special attention to the ability vector, given that prior ability represents an important competing explanation to schooling in the returns to education literature, and particularly so in the literature on selective schooling. In order to assess their correlation with our set of outcomes, we display the coefficients on the three indicators of ability in all our main regression tables, comparing them to the coefficients for school type. However, given that we are matching observations with school type as treatment, we note that the coefficients on the ability measures have no causal meaning.

[14]Kernel density estimates for local area characteristics are already largely similar for treated and control pupils prior to entropy balancing, as shown in Figures A2 and A3 in the Appendix (only 12 non-binary characteristics shown).

2.4.4 Heterogeneity and robustness checks

In addition to trying alternative matching strategies, we implement some additional specifications to further explore the relationship of interest and the robustness of our estimates. In a first check, we include interactions of the treatment and ability variables, in order to explore heterogeneity of treatment effect by cognitive and non-cognitive ability. We estimate

$$Y_i^j = \gamma_0^j + \gamma_1^j S_i^j + \gamma_2^j S_i^j \times C_i^{top50\%} + \gamma_3^j S_i^j \times NC_i^{top50\%} + \mathbf{X}'\gamma_4^j + \epsilon_i'^j \tag{2.6}$$

by OLS, where, for ease of notation, $\mathbf{X}$ is the vector of all individual characteristics as in Equation (2.5), including binary indicators for scoring above the median in the cognitive and non-cognitive skill distributions. The estimates of γ_2^j and γ_3^j reflect the additional effect of treatment S^j, for individuals in the top 50% of the ability distribution, compared to treated individuals in the bottom 50%, for whom the effect is simply γ_1^j, and to the base category of comprehensive pupils[15]. We further estimate similar models interacting the school type indicator with sex and high father SES.

A second additional specification distinguishes between comprehensive schools that were formerly grammar or secondary modern, versus comprehensive that are purpose-built. Given the NCDS cohort entered secondary school in 1969, only four years after Circular 10/65, we want to ensure that the effect estimate of school type is not confounded by comprehensive schools still transitioning from their grammar or secondary modern origin. Moreover, school type is retrieved in 1974, and therefore schools could potentially have transitioned to comprehensive status between when the NCDS cohort member started school and data retrieval. For analysis with the GC sample, we set grammar as the base category and two dummy treatment variable indicators: one for attending a comprehensive that is a former grammar CF_i, and one for attending a purpose-built comprehensive CB_i. A similar approach is then implemented with the SMC sample too. We estimate

$$Y_i^j = \delta_0^j + \delta_1^j CF_i + \delta_2^j CB_i + \mathbf{X}'\delta_3^j + \xi_i^j \tag{2.7}$$

by OLS, where $\mathbf{X}$ is again the vector of individual characteristics. We now distinguish between the effect of attending a comprehensive that could still present characteristics typical of a grammar (or secondary modern) (δ_1^j) and the effect of attending a purpose-built comprehensive (δ_2^j), compared to the base category of

[15] In a further alternative specification, we interacted treatment with ability quartiles, to increase model flexibility. This did not provide any additional information, and we therefore stick to the simpler interactions.

attending grammar (or secondary modern).

In a third check, we only include purely selective and purely comprehensive LEAs in the estimation of Equation (2.5). We define purely selective LEAs as those with no comprehensive places in 1971, as recorded in Comprehensive School Committee (1971), and purely comprehensive as those with 100% of places in comprehensive schools. Although the estimation sample shrinks significantly, the aim is to provide the analysis for areas where it can be ruled out that comprehensive schools experienced the same 'cream-skimming' of pupils and resources as secondary modern schools.

A further robustness check consists of implementing an IV strategy after balancing, as an alternative way to address endogeneity of treatment S_i. Under the assumptions of relevance and exclusion restriction of the instrument, two-stage least squares (2SLS) estimation methods can yield a consistent and unbiased estimate of treatment effect (Angrist and Pischke, 2009). As mentioned, the literature has used share of comprehensive schools in the LEA and political majority in the area as instruments (Basu et al., 2018; Galindo-Rueda and Vignoles, 2005). We use percentage of 13-year-old pupils attending comprehensive school in each LEA in 1971, corresponding to the NCDS cohort, as IV to instrument school type[16]. The variable was retrieved from an external data source, the Comprehensive School Committee 1971 Journal, and linked to NCDS by LEA of individual.

The instrument Z_i satisfies the relevance requirement $corr(Z_i, S_i) \neq 0$, since both grammar and secondary modern attendance are expected to be significantly and negatively correlated with the percentage of comprehensive school places in the LEA. Secondly, assuming exclusion restriction holds requires the instrument to only affect the outcome through its effect on treatment assignment, $cov(Z_i, \epsilon_i) = 0$, where ϵ_i is the error term in the main outcome equation, Equation (2.5). This seems plausible, although it is difficult to demonstrate. The main counter-argument would be that LEAs with higher values of the instrument (i.e. more comprehensive places) could be systematically different from ones with lower values. A good starting point for our IV is that LEA characteristics, such as county proportion of unemployed, council tenants, house owners and professional categories for household heads, are all largely comparable for individuals in the selective and non-selective system in our sample (see Table 2.2)[17].

The first stage of 2SLS estimation, the empirical counterpart of Equation (2.1), consists of the school assignment function, using percentage of comprehen-

[16]We prefer this instrument to percentage of schools, as it gives a finer measure of supply of school places. Data on LEA political control was not available to us.

[17]We further control for all LEA-level characteristics in the parametric specification of outcome regressions.

sive pupils in the individual's LEA as an instrument:

$$S_i^j = \eta_0^j + \eta_1^j Z_i + \mathbf{X}'\eta_2^j + \upsilon_i^j. \tag{2.8}$$

The second stage then uses the school type predicted in the first stage as a regressor for the outcome equation with α^{IVj} as the unbiased treatment effect estimate[18]:

$$Y_i^j = \beta_0^{IVj} + \alpha^{IVj} S_i^j + \mathbf{X}'\beta_1^{IVj} + \epsilon_i^{\prime\prime j} \tag{2.9}$$

We implement the IV strategy as a robustness check, but we prefer the matching + OLS estimates for the following reasons. First, under treatment exogeneity, OLS has superior finite sample properties to IV estimators, and smaller variance (Sargan, 1958). We conduct Durbin-Wu-Hausman tests of endogeneity of school type for all outcomes of interest, implementing the weights obtained via entropy balancing and including all available controls. For all our outcomes, the test, available in Appendix Table A2, fails to reject the null hypothesis of exogeneity of treatment. Second, as a rule of thumb, when confidence intervals for IV estimators contain OLS point estimates, it is advisable to use OLS, since this suggests the two estimators are not statistically different (Sargan, 1958; Young, 2019). We find this to be the case for all of our outcomes. On the basis of this evidence, we keep matching + OLS as our main empirical strategy (Mackinnon and Davidson, 2003). Our hypothesis is that by implementing the balancing algorithm on some key covariates and by including a rich set of control variables, we are able to control for some of the main confounders in the relationship between type of school and outcomes. We are however aware that credibility of our analysis relies on CIA validity, and we keep this into mind when interpreting our results.

The final robustness check follows the placebo test procedure implemented by Manning and Pischke (2006). Their procedure, detailed in the Appendix, consists of estimating the effect of type of secondary school for both pre- and post-secondary school maths test scores, at age 11 and 16 respectively. It is a placebo test because 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 is unable to prevent bias.

[18]Two Stage Residual Inclusion (2SRI) methods, allowing for non-linear models in either the first or second stage or both, were also explored as an alternative, but not included. See Terza et al. (2008) for more details on these methods.

2.5 Results

2.5.1 Characteristics by type of school

Table 2.1: Descriptive statistics of school characteristics by secondary school attended.

	Grammar		**Comprehensive**		**Sec. modern**	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Pupil composition						
Number of pupils	658.57	191.52	1107.76	377.94	670.33	283.41
Single sex (%)	0.68	0.47	0.13	0.34	0.26	0.44
Girls studying towards GCEs (%)	0.62	0.31	0.11	0.15	0.03	0.07
Boys studying towards GCEs (%)	0.66	0.33	0.11	0.15	0.02	0.06
>50% fathers in non-manual job	0.68	0.47	0.19	0.39	0.14	0.35
School characteristics						
Pupil-teacher ratio	16.05	1.55	17.13	2.07	18.19	2.30
Teachers left last year (%)	0.13	0.07	0.15	0.09	0.15	0.10
Parent-teacher association (%)	0.78	0.42	0.74	0.44	0.52	0.50
Lacks library (%)	0.25	0.43	0.21	0.41	0.23	0.42
Lacks science labs (%)	0.23	0.42	0.19	0.39	0.35	0.48
Lacks sport facilities (%)	0.32	0.47	0.34	0.47	0.38	0.49
Observations	1040		4663		1991	

Source: NCDS wave 3.

Table 2.1 summarises school characteristics by type for NCDS participants, showing higher peer ability and better resources for grammar schools among the three types. Grammar schools are on average smaller and more likely to be single sex. They display the highest percentage of pupils studying towards GCEs (General Certificate of Education, equivalent to A-levels[19]) and a higher probability of having more than 50% of fathers in non-manual occupations. In terms of resources, grammar schools have the lowest pupil teacher ratio of all and a lower teacher turnover. Higher proportions of grammar schools have a parent-teacher association and sports facilities, while secondary modern schools are more likely to lack science labs and sports facilities than other schools. Comprehensive are the least likely to lack a library and science labs.

Average individual characteristics prior to starting secondary school, displayed in Table 2.2, also present some differences by type of school. The largest gap is generally observed between grammar pupils and the other two groups, while average traits for comprehensive and secondary modern pupils are more similar. Future grammar, comprehensive and secondary modern pupils differ most notably in the three measures of ability. Grammar pupils present highest cognitive and non-cognitive abilities, followed by comprehensive and then by secondary modern pupils. On average, grammar pupils are also more advantaged in terms of socio-

[19] A-levels are the highest academic qualification that can be achieved in secondary school, corresponding to the US High School Diploma, the French *baccalauréat* and the German *Abitur*.

economic background. Average local area characteristics are very similar across the three groups, somewhat reassuringly for the identification of an unbiased treatment effect. The only notable exception, as expected, is the proposed IV, percentage of comprehensive pupils as a share of total pupils in the individual's LEA. This is highest for comprehensive pupils, compared to the rest of the sample.

Table 2.3 summarises all outcome variables used in the analysis. On average, grammar pupils display higher well-being, better labour market outcomes and better health, while secondary modern students fare worst out of the three groups. The only exception is life satisfaction, where comprehensive pupils score highest and grammar pupils score lowest.

2.5.2 Entropy balancing

We summarise entropy balancing results in Table 2.4, showing the first three moments of the five key covariates of interest before and after balancing, separately for the GC and SMC samples. The leftmost three columns in both the top and bottom panels show mean, variance and skewness for the treated group, while the three central columns show mean, variance and skewness for the unbalanced comprehensive sample. The last three columns on the right show weighted moments for control individuals from the comprehensive sample, using the entropy balancing weights. In both cases, the weighting procedure achieves almost perfect balance on mean, variance and skewness of key covariates, so that in the control group these are similar to the respective treated group. The pairwise interactions between covariates are not shown, but close 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. Figure 2.4 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 group, thus strengthening credibility of comprehensive pupils as counterfactual groups for our two separate treatment effects.

2.5.3 Selective schooling and long-term outcomes

Tables 2.5 to 2.7 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 marginal effects are shown. Although we only show the coefficients on the treatment and ability

Table 2.2: Descriptive statistics of covariates by secondary school attended.

	Grammar		Comprehensive		Secondary modern	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Ability						
Cognitive skills age 7	0.76	0.10	0.61	0.16	0.59	0.15
Non-cognitive skills age 11	0.94	0.08	0.88	0.12	0.86	0.13
Relative cognitive ability age 11	0.79	0.15	0.50	0.29	0.37	0.22
Background characteristics						
Female	0.55	0.50	0.48	0.50	0.49	0.50
Whether first born	0.36	0.48	0.31	0.46	0.30	0.46
Born in Wales	0.03	0.18	0.09	0.28	0.03	0.16
Not white	0.02	0.14	0.04	0.20	0.04	0.20
Two or more siblings	0.65	0.48	0.73	0.45	0.75	0.43
Twin or triplet	0.01	0.11	0.02	0.15	0.03	0.17
No mother	0.00	0.06	0.01	0.08	0.01	0.09
No father	0.03	0.16	0.04	0.20	0.04	0.20
Family SES						
Mother interest in child education	2.70	0.77	2.02	1.03	1.88	1.03
Father's SES high/middle-high	0.32	0.47	0.13	0.34	0.11	0.31
Father unemployed	0.01	0.09	0.03	0.17	0.04	0.18
Father job skilled/professional	0.54	0.50	0.47	0.50	0.47	0.50
Council housing	0.19	0.39	0.39	0.49	0.39	0.49
Free school meals	0.03	0.16	0.09	0.29	0.10	0.30
Health endowment						
Mother smoke when pregnant	1.37	0.78	1.59	0.92	1.60	0.93
Child morbidity index	0.06	0.03	0.06	0.04	0.06	0.04
Chronic condition in the family	0.11	0.31	0.15	0.36	0.15	0.36
Child in primary school						
Unhappy at school	0.03	0.18	0.07	0.26	0.07	0.26
Independent primary school	0.04	0.19	0.01	0.10	0.01	0.09
Child plans to study after school	0.43	0.50	0.23	0.42	0.17	0.37
LEA (1971 Census)						
% comprehensive pupils in LEA	0.29	0.25	0.52	0.32	0.24	0.21
County level % unemployed male	0.04	0.02	0.04	0.02	0.04	0.02
— council housing	0.28	0.08	0.29	0.08	0.28	0.08
— owner-occupiers	0.49	0.16	0.48	0.14	0.52	0.11
— manufacturing employee	0.34	0.12	0.36	0.11	0.36	0.10
— agriculture employee	0.02	0.04	0.02	0.03	0.02	0.03
— lone parent families	0.09	0.02	0.10	0.02	0.09	0.02
— UK born men	0.91	0.06	0.91	0.06	0.92	0.05
— professional/managerial HOH	0.18	0.08	0.16	0.07	0.16	0.06
— non manual HOH	0.22	0.07	0.21	0.06	0.20	0.05
— skilled manual HOH	0.27	0.09	0.28	0.08	0.29	0.07
— semi-skilled manual HOH	0.11	0.04	0.12	0.04	0.12	0.03
— non-skilled manual HOH	0.07	0.02	0.07	0.02	0.07	0.02
County borough	0.26	0.44	0.34	0.47	0.27	0.44
London borough	0.11	0.31	0.10	0.29	0.04	0.20
Observations	1040		4663		1991	

The three ability variables are bound between 0 and 1. Mother interest in child education is on a scale from 1-Little interest to 4-Over concerned. Maternal smoking during pregnancy is on a scale from 1-Non smoker to 4-Heavy smoker.

Table 2.3: Descriptive statistics of outcomes by secondary school attended.

	Grammar		**Comprehensive**		**Secondary modern**			
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Min	Max
Well-being measures								
School aspirations age 16 (dummy)	0.68	0.47	0.28	0.45	0.18	0.39	0	1
Work aspirations age 16 (dummy)	0.93	0.26	0.79	0.41	0.76	0.43	0	1
Life satisfaction age 33 (PCA)	-0.04	1.40	0.04	1.44	-0.02	1.49	-8	2
Self-efficacy age 33 (PCA)	0.21	1.17	-0.02	1.34	-0.06	1.34	-5	1
Positive about job age 33 (PCA)	0.37	1.20	-0.03	1.41	-0.15	1.44	-5	2
Contact with police age 45 (dummy)	0.14	0.35	0.18	0.38	0.18	0.38		
Ever tried illegal drugs age 45	0.19	0.39	0.17	0.38	0.17	0.37	0	1
Labour outcomes								
Hourly wage at 33	9.33	12.39	7.19	12.52	6.42	10.86	0	357
Employed at 33	0.84	0.37	0.79	0.40	0.80	0.40	0	1
Hourly wage at 50	22.30	30.09	16.33	12.91	15.16	10.97	0	462
Employed at 50	0.92	0.27	0.86	0.35	0.85	0.36	0	1
Survey health measures								
Excellent or very good SAH age 50	0.62	0.49	0.52	0.50	0.48	0.50	0	1
Low malaise age 50	0.81	0.39	0.77	0.42	0.76	0.43	0	1
Mental ill-health score age 45	3.02	4.17	3.40	4.68	3.40	4.63	0	30
Biometric health measures								
BMI measured age 45	26.41	4.61	27.56	4.88	27.67	5.16	17	64
Cholesterol ratio mmo/L age 45	3.80	1.15	3.97	1.17	4.07	1.18	2	12
Triglyceride mmo/L age 45	1.88	1.46	2.06	1.61	2.15	1.71	0	27
Fibrinogen g/L age 45	2.88	0.56	2.98	0.63	3.00	0.62	1	7
C reactive protein mg/L age 45	1.84	3.35	2.27	4.93	2.26	4.26	0	152
Observations	1040		4663		1991			

Source: NCDS. For more details on the variables constructed by PCA, see Appendix. For the wage outcome, we excluded from the analysis 13 individuals with weekly income above £10,000. Healthy ranges for the biometric markers are as follows: <25 for BMI, <5 for cholesterol ratio, <1.7 for triglycerides, 1.9-4.3 for fibrinogen and <5 for CRP (Fuggle, 2018).

Table 2.4: Pre- and post-matching moments of key covariates.

	Grammar N=1040			Raw comprehensive N=4663			Balanced comprehensive N=4663		
	Mean	Variance	Skewness	Mean	Variance	Skewness	Weighted m.	Weighted v.	Weighted s.
Cognitive skills	0.763	0.010	-0.467	0.618	0.025	-0.404	0.763	0.010	-0.469
Non-cognitive skills	0.940	0.006	-2.288	0.882	0.015	-1.513	0.940	0.006	-2.286
Relative cognitive score	0.795	0.021	-0.944	0.507	0.082	-0.029	0.795	0.021	-0.946
Mother's interest in edu	2.697	0.585	-1.843	2.027	1.057	-0.490	2.697	0.585	-1.842
High father's SES dummy	0.822	0.146	-1.685	0.692	0.213	-0.831	0.822	0.146	-1.683
	Secondary modern N=1991			**Raw comprehensive N=4663**			**Balanced comprehensive N=4663**		
	Mean	Variance	Skewness	Mean	Variance	Skewness	Weighted m.	Weighted v.	Weighted s.
Cognitive skills	0.590	0.022	-0.308	0.618	0.025	-0.404	0.590	0.022	-0.308
Non-cognitive skills	0.867	0.016	-1.363	0.882	0.015	-1.513	0.867	0.016	-1.362
Relative cognitive score	0.376	0.047	0.363	0.507	0.082	-0.029	0.376	0.047	0.364
Mother's interest in edu	1.908	1.065	-0.317	2.027	1.057	-0.490	1.908	1.065	-0.317
High father's SES dummy	0.671	0.221	-0.728	0.692	0.213	-0.831	0.671	0.221	-0.728

The top panel refers to the GC sample, while the bottom panel to the SMC sample. Mean, variance and skewness of the pairwise interactions of the five covariates listed are also balanced (not shown).

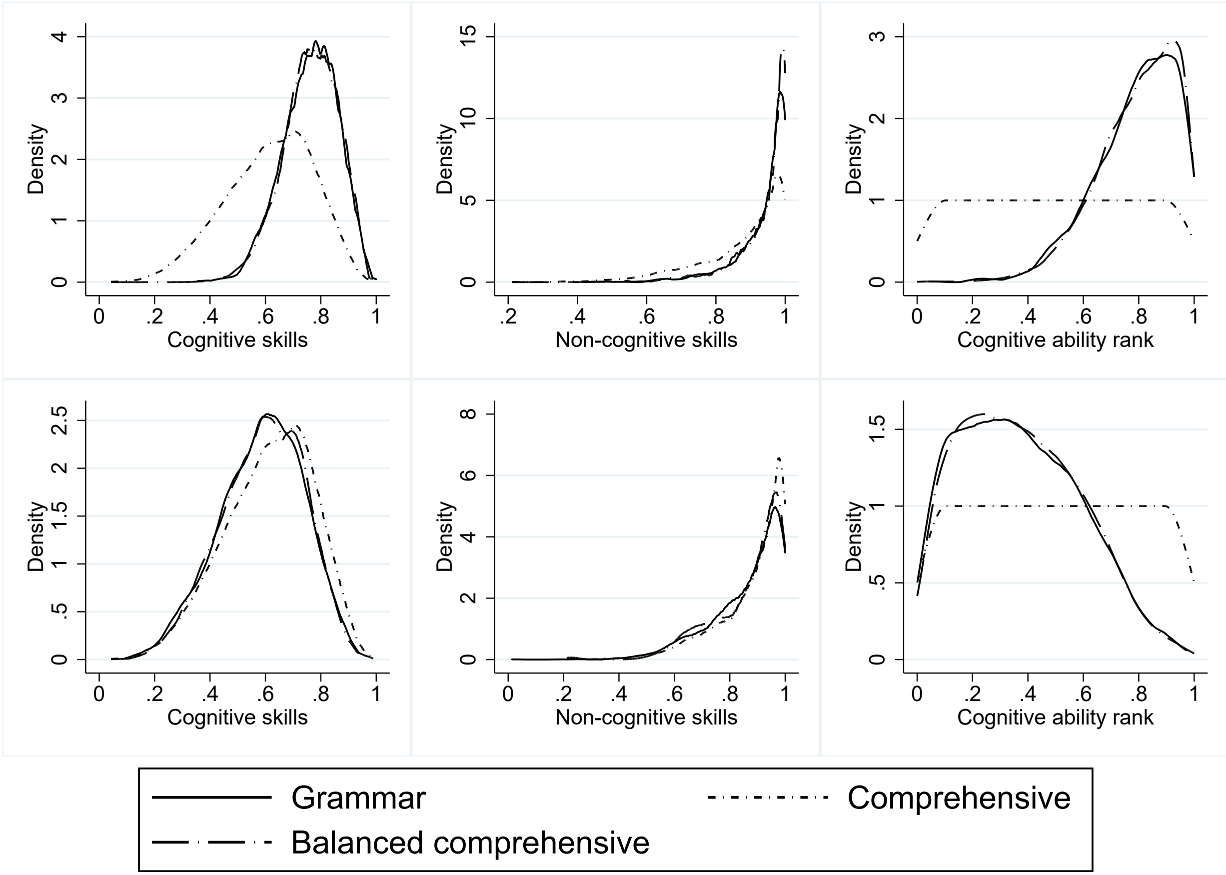


Figure 2.4: Kernel density estimates for three ability measures, for the GC sample (top row) and the SMC sample (bottom row). Mother's interest in child education and father's SES are not shown, as they are discrete variables. The dashed line (large dash) illustrates density kernels for comprehensive pupils, balanced with the weights obtained via entropy matching so that they are more comparable to treated individuals.

variables, all models are estimated controlling for all covariates described in Table 2.2.

Table 2.5 shows that, for the high cognitive ability sample, attending a grammar school significantly increases school aspirations, the probability that individuals intend to stay at school beyond 16, by approximately 13 percentage points (pp), compared to comprehensive attendance. At the same time, grammar attendance decreases adult life satisfaction by 0.13 standard deviations (SD). For the lower cognitive ability sample in the lower panel, secondary modern attendance decreases the intention to stay at school beyond minimum leaving age (4 pp), while it increases self-efficacy at 33 (0.08 SD), compared to comprehensive. In Table 2.6, showing estimation results for labour market outcomes, grammar is significant at 10% for both age 33 and age 50 log-transformed wages, raising average hourly wage by 6 pp and 9 pp, compared to attending comprehensive. Grammar also increases the probability of being employed at 33 by 3 pp, compared to similarly able comprehensive pupils. Secondary modern attendance, in the lower panel, increases average wage at 50 by roughly 8 pp, and the probability of being employed at 33 by 3 pp. Table 2.7 displays results for health outcomes. Grammar attendance is only significantly related to BMI, decreasing it by 0.1 SD, compared to comprehensive, while secondary modern, in the bottom panel, is only significantly and positively related to cholesterol ratio (p-value<0.1).

The ability variables, on the other hand, display a significant association with most outcomes. As mentioned in Section 2.4.3, while lacking causal meaning, these coefficients can guide us in assessing how ability compares with school type as a competing explanation for several future outcomes. In Table 2.5, for both the GC and SMC sample, higher cognitive ability at age 7 is linked to higher self-efficacy in adulthood (0.4-0.6 SD for a one-unit increase in cognitive ability[20]). Age 7 cognitive skills are linked to employment outcomes only in the low ability sample (up to 18 pp increase in age 33 wages). Conversely, they display a significant link with lower biomarkers (BMI, cholesterol ratio and triglycerides), indicative of better health, in the high ability sample only.

In both samples, age 11 non-cognitive skills are linked with higher life satisfaction and self-efficacy (0.6-1.2 SD and 0.4-1.1 SD respectively), and lower probability of drug use (24 to 30 pp). The association is stronger in magnitude for the GC sample, of higher cognitive ability. Non-cognitive skills are also significantly associated with the probability of aspiring to personal and intellectual growth at work (13 pp), job positivity (0.5 SD) and a lower probability of committing

[20]A one-unit increase in either the cognitive skills, non-cognitive skills or relative cognitive ability variables corresponds to a move from the bottom to the top of the distribution, since the three indices range between 0 and 1. In this section, all coefficients express the change in outcome associated to a one-unit move in the ability indices.

crime (24 pp) in the SMC sample only. Table 2.6 shows that non-cognitive ability is also positively and significantly associated with wages (up to 75 pp increase), and with the probability of employment (up to 18 pp). Further, non-cognitive skills are significantly related to better health in the SMC sample, with a 35 pp increase in the probabilities of scoring high self-assessed health and low malaise at age 50, and a reduction of 0.6 SD in mental health problems at 45. Compared to the GC sample, this result may indicate that higher non-cognitive skills have a protective role for the health of pupils of lower average cognitive ability.

Finally, age 11 cognitive ability rank is significantly and positively linked to positive school (35 to 70 pp increase) and work aspirations (18 to 41 pp), and job positivity (0.6-0.8 SD). Relative cognitive ability is significantly linked to wages, and coefficients are large, indicating increases of up to 80 pp in the GC sample and up to 44 pp in the SMC sample. Moving from the lowest to the highest rank of cognitive ability is also linked to a decrease in C-Reactive protein (0.5 SD) and fibrinogen levels (0.4 SD) in the GC sample. In the SMC sample, higher cognitive rank is significantly associated with the probabilities of scoring high self-assessed health (15 pp increase), low malaise (12 pp increase), and lower fibrinogen levels (0.4 SD).

2.5.4 Heterogeneous effects

Interacting treatment with high levels of cognitive and non-cognitive skills, as shown in Equation (2.6), did not add any further insight to our main message, as illustrated in Tables A3-A5 and also in figures A4-A9 in the Appendix, plotting average treatment effect estimated at each point of the ability distribution, for the three ability measures considered. We only note that grammar appears to have a larger positive effect on school aspirations and wages for pupils in the bottom half of the cognitive ability distribution for the GC sample (Tables A3-A4). Further interactions with sex, in Appendix Tables A6-A8, indicate that the grammar advantage for school aspirations was more prominent for boys, while the secondary modern disadvantage for this outcome is only significant for girls. For girls, attending grammar is also significantly linked to lower self-efficacy in adulthood. As for employment outcomes, Table A8 shows no significant difference by sex in the grammar coefficient, while the secondary modern advantage in terms of employment and wages appears to be driven by boys (although precision in these coefficients varies). More differences by sex can be observed in Table A8. The grammar coefficient shows a positive association with low malaise scores for boys, and negative for girls. Conversely, grammar attendance is linked to higher CRP levels (increasing in bad health) for boys and lower for girls. Interacting

Table 2.5: Selective schooling and well-being outcomes.

	School asp.	Work asp.	Life sat.	Self-eff.	Job positiv.	Crime	Drugs
Grammar vs comprehensive (high ability)							
Grammar	0.1259***	0.0130	-0.1307**	-0.0535	-0.0014	-0.0142	-0.0103
	(0.0165)	(0.0111)	(0.0460)	(0.0371)	(0.0474)	(0.0171)	(0.0183)
Cognitive skills	0.1375	-0.0625	-0.0561	0.5919**	0.2390	0.0706	-0.0051
	(0.0943)	(0.0557)	(0.2744)	(0.2164)	(0.2448)	(0.0971)	(0.0911)
Non-cognitive skills	0.1692	0.0615	1.1847**	1.1133***	0.0421	-0.0450	-0.3070**
	(0.1185)	(0.0663)	(0.3625)	(0.3228)	(0.2596)	(0.0964)	(0.0983)
Relative cogn. ability	0.6918***	0.1832***	-0.1145	0.3567*	0.8344***	-0.0783	0.1491*
	(0.0623)	(0.0420)	(0.1602)	(0.1622)	(0.1550)	(0.0587)	(0.0608)
Observations	4197	4156	3131	3083	3145	3277	3279
F statistic			5.3818	7.7965	14.2346		
χ^2 statistic	305.38	52.28				63.44	65.97
Secondary modern vs comprehensive (low ability)							
Secondary modern	-0.0420**	0.0179	-0.0069	0.0789*	0.0258	-0.0207	-0.0066
	(0.0129)	(0.0142)	(0.0439)	(0.0362)	(0.0389)	(0.0131)	(0.0142)
Cognitive skills	0.0393	0.0315	0.2013	0.3840+	0.2571+	0.0840	0.1064*
	(0.0520)	(0.0569)	(0.1581)	(0.1947)	(0.1531)	(0.0551)	(0.0511)
Non-cognitive skills	0.1413*	0.1341*	0.5838***	0.3905*	0.4602**	-0.1255*	-0.2386***
	(0.0553)	(0.0570)	(0.1631)	(0.1603)	(0.1426)	(0.0543)	(0.0572)
Relative cogn. ability	0.3449***	0.4142***	-0.2027*	0.2248*	0.6090***	-0.0432	0.0329
	(0.0360)	(0.0413)	(0.0957)	(0.1095)	(0.1164)	(0.0429)	(0.0396)
Observations	4872	4818	3588	3535	3597	3777	3779
F statistic			5.7861	9.7086	20.4473		
χ^2 statistic	282.59	275.40				206.80	83.91

+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001. Standard errors clustered at LEA level in parentheses. All continuous outcomes are standardised. Binary outcomes are estimated via probit models, for which marginal effects are displayed. All control variables are included.

Table 2.6: Selective schooling and labour market outcomes.

	Log hourly wage 33	**Employed at 33**	**Log hourly wage 50**	**Employed at 50**
Grammar vs comprehensive (high ability)				
Grammar	0.0588+	0.0321+	0.0879+	0.0199
	(0.0308)	(0.0164)	(0.0509)	(0.0144)
Cognitive skills	0.0082	0.0783	-0.0131	0.0676
	(0.1666)	(0.0970)	(0.2810)	(0.0856)
Non-cognitive skills	0.4218*	0.0841	0.7541*	0.0963
	(0.2046)	(0.1194)	(0.2990)	(0.0766)
Relative cogn. ability	0.6533***	0.0585	0.7986***	-0.0165
	(0.1181)	(0.0633)	(0.1563)	(0.0459)
Observations	2460	3323	1551	2852
F statistic	27.2839		9.6002	
χ^2 statistic		189.20		47.44
Secondary modern vs comprehensive (low ability)				
Secondary modern	0.0474	0.0338*	0.0827*	-0.0205
	(0.0324)	(0.0154)	(0.0388)	(0.0132)
Cognitive skills	0.1868*	0.0422	0.0552	0.1155*
	(0.0835)	(0.0646)	(0.1315)	(0.0550)
Non-cognitive skills	0.2504*	0.1375**	-0.0229	0.1818***
	(0.1091)	(0.0495)	(0.1808)	(0.0550)
Relative cogn. ability	0.3710***	0.0566	0.4450***	0.1102**
	(0.0642)	(0.0421)	(0.0795)	(0.0376)
Observations	2766	3821	1689	3230
F statistic	29.7419		20.8571	
χ^2 statistic		268.53		110.05

+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001. Standard errors clustered at LEA level in parentheses. Binary outcomes are estimated via probit models, for which marginal effects are displayed. All control variables are included.

Table 2.7: Selective schooling and health outcomes.

	High SAH	Low malaise	Mental ill-health	BMI	Chol ratio	Triglycerides	CRP	Fibrinogen
Grammar vs comprehensive (high ability)								
Grammar	-0.0055	0.0170	0.0122	-0.1056*	0.0385	-0.0069	0.0031	0.0133
	(0.0301)	(0.0210)	(0.0486)	(0.0528)	(0.0501)	(0.0476)	(0.0449)	(0.0587)
Cognitive skills	0.2267+	0.0699	-0.1821	-0.6314**	-0.5410+	-0.5765+	-0.1312	-0.0597
	(0.1257)	(0.1107)	(0.2928)	(0.2337)	(0.3019)	(0.2984)	(0.2029)	(0.2834)
Non-cognitive skills	0.2213	0.0946	-0.6169*	-0.4083	-0.3525	-0.2488	-0.2157	-0.5401+
	(0.1558)	(0.1233)	(0.2917)	(0.2819)	(0.3540)	(0.3497)	(0.2149)	(0.3140)
Relative cogn. ability	0.0948	0.1047	0.2404	0.0119	-0.2163	-0.0790	-0.4670*	-0.4220+
	(0.1144)	(0.0764)	(0.1821)	(0.2003)	(0.1811)	(0.2501)	(0.1910)	(0.2408)
Observations	2875	2854	2805	2759	2327	2333	2302	2295
F statistic			9.4540	4.9628	56.5408	47.2199	3.5059	3.8530
χ^2 statistic	52.8400	56.4290						
Secondary modern vs comprehensive (low ability)								
Secondary modern	-0.0034	0.0060	-0.0302	0.0355	0.0674+	0.0241	-0.0231	-0.0281
	(0.0237)	(0.0191)	(0.0372)	(0.0479)	(0.0405)	(0.0447)	(0.0477)	(0.0426)
Cognitive skills	0.0623	0.0398	-0.1485	-0.0862	-0.0261	-0.0826	-0.2524	-0.0479
	(0.0831)	(0.0738)	(0.1620)	(0.1881)	(0.1991)	(0.2214)	(0.2128)	(0.1731)
Non-cognitive skills	0.3468***	0.3456***	-0.6103***	0.1171	-0.2685	-0.2160	0.1938	0.0904
	(0.0857)	(0.0808)	(0.1639)	(0.1820)	(0.2452)	(0.2208)	(0.2016)	(0.1957)
Relative cogn. ability	0.1516**	0.1223**	-0.1603	-0.1411	-0.0989	-0.0803	-0.1599	-0.4420***
	(0.0540)	(0.0466)	(0.1217)	(0.1315)	(0.1302)	(0.1096)	(0.1463)	(0.1183)
Observations	3250	3224	3203	3145	2665	2669	2634	2629
F statistic			6.0568	6.2434	17.0289	25.9164	3.9511	6.0538
χ^2 statistic	109.2700	96.6975						

+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001. Standard errors clustered at LEA level in parentheses. All continuous outcomes are standardised. Binary outcomes are estimated via probit models, for which marginal effects are displayed. All control variables are included.

treatment with SES (not shown) did not offer additional information.

Secondly, distinguishing between comprehensive schools by origin did not produce significantly different results from our main specification. We leave results in Appendix Tables A9-A11. The main new insight is that the wage results are driven by differences between grammar/secondary modern schools and purpose-built comprehensive ones. Conversely, former grammar and secondary modern converted into comprehensive did not produce significantly different wage outcomes in those early years of the transition.

Finally, when including in the estimation procedure only LEAs that were purely selective or purely non-selective, sample sizes shrunk significantly, but the main results still apply (see Appendix Tables A12-A14). Grammar attendance is linked to an increase in the probability of wanting to stay in school beyond minimum leaving age, and to better employment outcomes (although with lower significance levels), as well as lower BMI. In this specification, coefficients are larger and grammar attendance is also linked to lower CRP and fibrinogen levels, indicative of good health. For secondary modern attendance, the link with school aspirations is negative as in the main specification, although not significant, and the positive links with self-efficacy and employment at 33 are confirmed.

2.5.5 Instrumental variable estimates

The first stage of the IV specification, used as a robustness check for our main results, shows a significant and negative correlation between the instrument, percentage of comprehensive pupils in the individual's LEA, and the treatment variable in both samples (see Appendix Table A15). The partial F-test is always greater than 10, which by rule of thumb increases the confidence that the instrument of choice is not weak[21] (Staiger and Stock, 1997). Tables A16 to A18 in the Appendix display results for the second stage of 2SLS models. Standard errors are in general at least twice as large as OLS standard errors. Similarly to OLS, the 2SLS estimator suggests that grammar attendance significantly increases school aspirations, i.e. the likelihood of wanting to stay in school beyond 16, by 11 pp, while secondary modern attendance decreases it by almost 6 pp. The coefficient of grammar on life satisfaction is negative but not significant, due to a larger standard error. The largest difference is observed with labour outcomes, where the coefficients of grammar are larger for age 33 outcomes and insignificant for age 50 wage, while the coefficients for secondary modern are insignificant. Still, most of the point estimates across all outcomes are similar to OLS ones in sign

[21] 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).

and magnitude, and when the sign changes, it is always for near-zero estimates. Moreover, 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[22].

2.5.6 Falsification tests

As a further check, we conduct placebo procedures in the same spirit of Manning and Pischke's falsification test, in order to support credibility of our empirical strategy. There are some key differences 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, we include the weights obtained by entropy balancing and the set of control variables used in our main specification; third, by varying the balancing algorithm, we try to assess the extent of the potential bias due to unobservables or misspecification. Following the original paper, maths test scores are converted to a scale from 0 to 100, so that coefficients are more easily interpreted. The first four columns in Table 2.8 display results for age 11 maths scores, while the last four do the same for age 16 scores. Results for age 11 scores in columns (1) and (2) for the GC sample confirm what was found by the original authors. Comprehensive attendance, used as treatment for both groups for comparability with the original test, is a significant and negative predictor of age 11 maths scores when compared to grammar, even after balancing. However, the magnitude of the coefficient is halved after balancing, which suggests the matching procedure is working in the right direction.

One hypothesis is that the residual 8.76 percentage point difference in age 11 outcomes between grammar and comprehensive pupils is due to the the coaching effects mentioned in Section 2.4.2. Primary schools in selective areas were likely to tutor their pupils in preparation for the 11-plus, and even short-term coaching has been shown to have large positive effects on performance on this type of test, warranting a relatively large size of the coaching effect (Bunting and Mooney, 2001). For the SMC sample, comprehensive attendance is a significant predictor of maths scores at age 11. The coefficient is positive in the unmatched sample (column (1)), and negative in the matched one (column (2)). In this case, entropy balancing potentially eliminates differences in observable characteristics that introduce positive bias in the comprehensive coefficient in the unmatched

[22]Two Stage Residual Inclusion (2SRI) results, available on request, are also close to OLS results. 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.

Table 2.8: Placebo age 11 regressions and age 16 regressions.

	Age 11 maths scores				Age 16 maths scores			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Unmatched	Matched	Matched	Matched	Unmatched	Matched	Matched	Matched
Grammar vs comprehensive (high ability)								
Comprehensive	-0.1551***	-0.0876***	-0.0105	-0.0123	-0.0887***	-0.0656***	-0.0664***	-0.0616***
	(0.0083)	(0.0074)	(0.0081)	(0.0081)	(0.0075)	(0.0072)	(0.0071)	(0.0072)
Cognitive skills 7	0.8174***	0.5230***	0.5173***	0.4836***				
	(0.0217)	(0.0331)	(0.0307)	(0.0333)				
Cognitive skills 11					0.7535***	0.8565***	0.8926***	0.8917***
					(0.0152)	(0.0302)	(0.0302)	(0.0319)
Non-cognitive skills	0.2712***	0.0601	0.0651	0.0794+	0.1011***	0.2517***	0.2599***	0.2527***
	(0.0212)	(0.0412)	(0.0473)	(0.0460)	(0.0213)	(0.0456)	(0.0486)	(0.0472)
Matched on age 7 cognitive skills	No	Yes	No	Yes	No	Yes	No	Yes
Matched on age 11 cognitive skills	No	No	Yes	Yes	No	No	Yes	Yes
Observations	4166	4166	4166	4166	4166	4166	4166	4166
F statistic	255.0254	25.7752	28.6748	19.4825	367.1590	84.9778	93.3781	92.6273
Secondary modern vs comprehensive (low ability)								
Comprehensive	0.0363***	-0.0427***	0.0073	0.0118+	0.0173**	0.0054	0.0098+	0.0087
	(0.0064)	(0.0058)	(0.0060)	(0.0062)	(0.0055)	(0.0054)	(0.0053)	(0.0053)
Cognitive skills 7	0.8009***	0.6195***	0.6917***	0.6755***				
	(0.0177)	(0.0181)	(0.0174)	(0.0172)				
Cognitive skills 11					0.7185***	0.6108***	0.6333***	0.6304***
					(0.0138)	(0.0171)	(0.0152)	(0.0152)
Non-cognitive skills	0.2667***	0.2321***	0.2249***	0.2460***	0.0855***	0.0840***	0.0943***	0.0880***
	(0.0207)	(0.0215)	(0.0231)	(0.0233)	(0.0171)	(0.0199)	(0.0207)	(0.0208)
Matched on age 7 cognitive skills	No	Yes	No	Yes	No	Yes	No	Yes
Matched on age 11 cognitive skills	No	No	Yes	Yes	No	No	Yes	Yes
Observations	4847	4847	4847	4847	4847	4847	4847	4847
F statistic	231.5945	86.1400	144.3652	145.5350	179.1259	67.9224	93.7780	94.5217

+ $p < 0.1$ *** $p < 0.001$ Standard errors clustered at LEA level in parentheses. All control variables included.

sample. The residual 4.27 percentage points could also be due to coaching effects. Although it seems large, it is half the coefficient of comprehensive in the GC sample, which is in line with average cognitive ability of future secondary modern pupils being lower than that of grammar pupils[23].

An alternative explanation to coaching effects could be that the coefficients in column (2) reflect pre-treatment differences that the matching + OLS strategy is not adequately accounting for. In this case, it is useful to assess whether eliminating these differences would affect treatment effect, in order to put bounds on the potential bias. To this end, in columns (3) and (7), we balance the samples on age 11 cognitive ability scores instead of age 7 scores. Additionally, in columns (4) and (8), we include both age 7 and 11 scores in the balancing algorithm. We expect both of these alternative balancing algorithms to eliminate differences in age 11 maths scores between treatment and control groups, thus making the comprehensive indicator insignificant, since age 11 maths scores are highly correlated to age 11 cognitive scores[24]. The aim is to check whether eliminating residual age 11 differences has any effect on the comprehensive coefficient for age 16 maths scores, to understand the importance of any potential bias due to unobservables for later outcomes.

Columns (5)-(8) in Table 2.8 show that the magnitude of the comprehensive coefficient declines in the GC sample, when moving from the unmatched to the matched sample, but remains fairly stable around 6 percentage points, even when artificially eliminating age 11 differences. Something similar can be observed for the SMC sample, where the coefficient is initially positive and significant at 1.73 percentage points (column (5)). This then decreases to less than 1 percentage point, becoming null, when matching, remaining stable across the different balancing algorithms (columns (6) to (8)). In both samples, eliminating potential unobservable differences at age 11 by matching on age 11 cognitive scores produces a variation in the age 16 coefficient of less than 0.005. This is compatible with the coefficients in column (2) picking up a large, although relatively short-lived, coaching effect that biases age 11 maths scores upwards for pupils in selective areas, which would then disappear by age 16[25]. Moreover, to the extent that coaching does not affect outcomes outside of test scores, the range of outcomes

[23]We cannot directly compare our coefficient with Manning and Pischke (2006), since our samples are different, but their estimates range between 3 and 8 pp, depending on the specification.

[24]Recall that age 11 ability score is constructed by PCA on maths, reading and general ability tests. As detailed in Section 2.4.2, we do not match on age 11 scores in the main specification, because these could be biased by coaching, while we are interested in matching individuals on their underlying 'true' ability, since that is potentially correlated with our main outcomes.

[25]In addition to this, age 7 outcomes are not significantly different between the balanced treatment and control groups.

presented in this paper will be exempt from this bias[26].

2.6 Discussion

When correcting for pre-treatment differences in pupil characteristics, type of school is a significant predictor of school aspirations and some labour market outcomes. Most well-being measures are not affected, with the exception of life satisfaction and self-efficacy, and in some specifications only. We find that in the short-term, pupils who attended grammar schools in the 1970s were 13 percentage points more likely to say they intended to stay at school beyond minimum leaving age, than comprehensive pupils of similar ability. This effect may translate into better educational qualifications and better employment prospects later in life, which is consistent with the evidence we present on wages (up to 9 points higher) and with a study on the earnings of more recent generations by Burgess et al. (2019). In spite of this, we also find that grammar lowers adult life satisfaction (by 0.13 standard deviations), compared to similarly able pupils in mixed-ability systems. Conversely, secondary modern pupils were about 5 percentage points less likely to want to stay at school beyond 16, maybe due to the vocational nature of the secondary modern curriculum.

While the link between school type and educational aspirations is of the expected sign, the positive effect of attending secondary modern on employment outcomes may be surprising. We advance some possible explanations as to why this is observed. First, better wages (up to 8 points higher) for secondary modern pupils may be linked to the fact they could learn a profession at school, compared to the general education given to low-ability pupils in comprehensive schools. Supporting this point, Appendix Figure A10 shows that, at age 33, the majority of former secondary modern and comprehensive NCDS pupils end up in skilled non-manual, skilled manual or partly skilled occupations. If secondary modern pupils finished school with more practical knowledge, it is reasonable that they would fare better in such professions. Additionally, this would be consistent with the significant and positive coefficient associated to secondary modern attendance for adult self-efficacy (0.08 SD). A second hypothesis relates to differences in labour market conditions in selective and non-selective areas at the time in which NCDS individuals left school. Although we have reduced variability in area-level characteristics via entropy balancing, it may be that we are unable to account for specific features indicating preparedness of the labour market

[26]When we carry out similar procedures with BMI at age 16 and 11 as dependent variables we find that comprehensive attendance is not significant for age 16 nor age 11 BMI, in either sample. Basu et al. (2018) conduct a similar check, taking age 11 child morbidity as main outcome of the placebo procedure.

for young school-leavers, such as the availability of apprenticeships. If this was the case, then it would be incorrect to attribute this effect to school type only. Third, characteristics common to schools in the selective school system, other than selection itself, could explain the outcomes. For instance, Table 2.1 shows that schools in the selective system are more likely to be single sex than comprehensive schools. The importance of this feature has been widely documented in the returns to education literature, and it could arguably also play a role in this context (Sullivan et al., 2011).

We further find that most of the health outcomes considered are not affected by school type, except for a link between grammar attendance and lower BMI, and one between secondary modern attendance and higher cholesterol ratio, although magnitudes are small, and significance levels vary by specification. Our findings corroborate previous literature on the long-term effects of selective schooling. 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. With respect to health effects in the English context, Basu et al. (2018) find no significant average effects of the transition to comprehensive schools for self-assessed health and smoking. We extend their result by adding evidence on biomarkers for risk of CVD and well-being outcomes, and by distinguishing between two treatment effects and two samples, in order to build the counterfactual control group more reliably. When exploring heterogeneity, Basu et al. (2018) find a negative effect of the move to mixed-ability schooling on depression for men with lower childhood non-cognitive skills. In our application however, we do not find that our simpler method to account for skills heterogeneity changes our main findings. We instead find some differences by sex, although generally to the effect that boys experience larger benefits from grammar schools, while girls experience larger disadvantages from secondary modern school.

Taken together, our evidence adds to the literature casting doubts over the effectiveness of selective schooling policies to improve long-term health and well-being outcomes. The significant differences shown in outcomes by school type in Table 2.3 mostly disappear once we take background into account, the only exceptions being educational aspirations, labour market outcomes and the life satisfaction measures mentioned. Yet, a note of caution is in order when interpreting the significance of these result for current policy: while grammar schools maintain their name and focus on academic education, the education system has evolved, and so have labour market conditions. Old secondary modern schools have either become comprehensive or shifted their focus towards more academic subjects, leaving vocational education to the remit of colleges offering National Vocational Qualifications (NVQs) and to company-based apprenticeships. Several new types of schools, among them faith schools, academies and free schools, have been created and compete with grammar schools for quality of teaching and pupil achievement. A prediction of the consequences of the current expansion policy will have to take these and other factors into account.

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 roll-out. 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, suggest that our methodology nets out some of the pre-treatment differences between compared groups. We argue that the residual placebo effect on age 11 maths scores is likely due to coaching effects. However, we additionally show that, even if the residual effect was due to unobservable differences at age 11, the bias is unlikely to carry over to later outcomes.

The validity of our analysis relies on the conditional independence assumption (CIA), assuming that, conditional on the observed variables, treatment is as good as randomly assigned. It is hardly plausible that all potential confounders are observed in practice, but we argue that we can still preserve unbiasedness if we assume that unobservables are sufficiently correlated with observed covariates, so that achieving balance on the latter also implies balance on unobserved confounders. If this assumption does not hold, then our OLS coefficients will be biased. The main candidates for unobservables that could enter both treatment and outcome equations, biasing the estimates, are underlying child ability, child propensity to schooling (including genetic factors), parental support and resources and peers' influence on the child. With exception for the latter, we have included proxies of these variables in our regressions. Given the expected effect of ability, propensity to schooling and parental support on outcomes is positive, we expect most of these unobservable traits to be positively as well as sufficiently correlated to the observed proxy covariates included in the matching and regression procedures. If these assumptions hold, we can then hypothesise that any residual bias could be driven mainly by genetic factors or peers' influence.

As a secondary result, we additionally note that childhood cognitive abilities are highly correlated with several of our 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). Similarly to Jones et al. (2011), we also find that childhood non-cognitive abilities are significantly linked with several long-term outcomes, and extend this result to our health, well-being and labour market outcomes. As shown in Figure 2.5, the association between all ability measures and outcomes is sizeable. The effect of school type is generally equivalent to moving by between 20 and 30 percentiles on the ability distribution. 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.

Building on this finding, a reason why we do not find a significant effect of selective schooling on health and well-being could be that secondary school type does not directly affect the channels leading to better adult well-being. In line with this, Jerrim and Sims (2018, 2019) find no association between attending grammar school and adolescent non-cognitive skills, for a cohort of British individuals born in 2000-2001. Cognitive and non-cognitive skills, as well as preferences determining our decisions, might then be shaped earlier on in childhood (Kautz et al., 2014). If it is true that skills and preferences 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. Further skill production could occur in adulthood via university attendance, career path, work and home environment and so on. Understanding the mechanisms and timing of skill production can then be a productive avenue for future research.

2.7 Conclusion

We add a piece of evidence to the debate on selective schools in England, by looking at the long-term health, well-being and labour market effects of attending a high- or low-ability school in a selective system, compared to a mixed-ability school in a non-selective system. We use data from NCDS, including individuals from both systems, allowing us to explore the effects of selective schooling at several points of these individuals' lives over time, in spite of the lack of a clear roll-out of the educational reform. We build our quasi-experimental framework by preprocessing the data through entropy balancing to obtain treatment and control groups with similar distributions and joint distributions of key covariates. We then use the balanced samples to implement OLS regression analysis and several robustness checks, including IV regressions, to strengthen the credibility of our results.

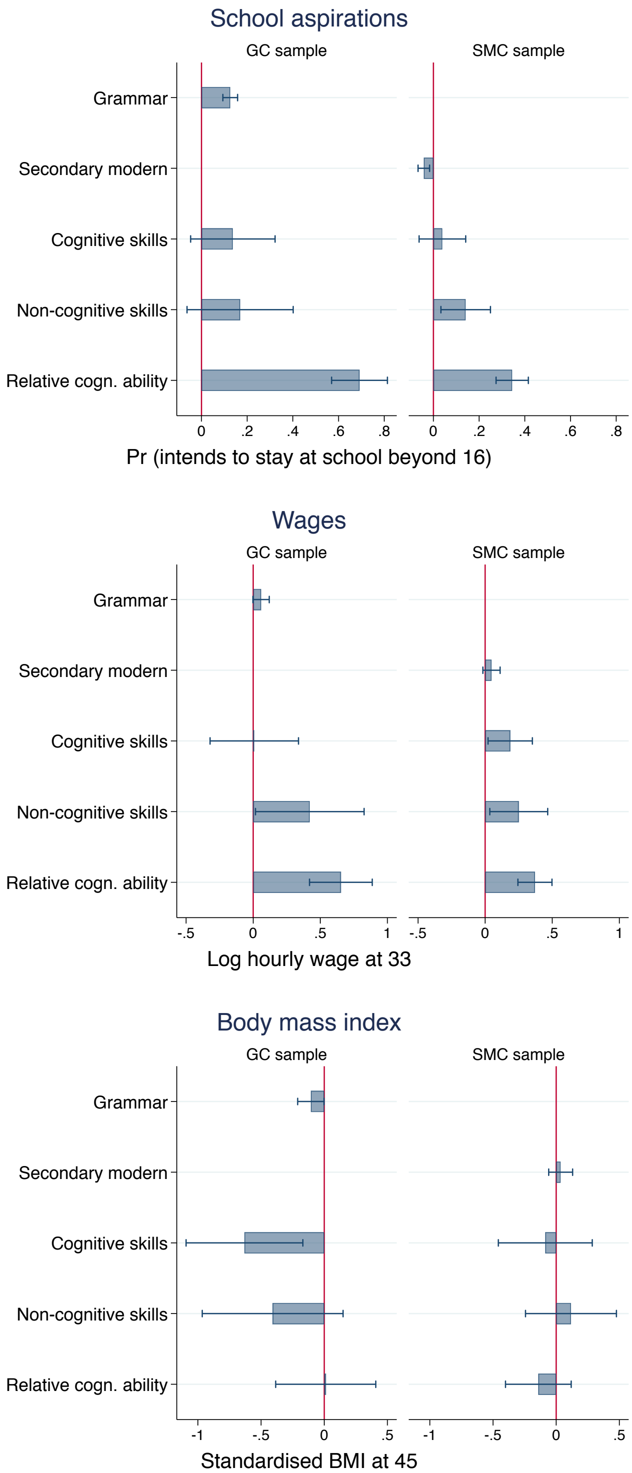


Figure 2.5: Coefficient size for school type and ability variables compared, for three outcomes. Estimates from Tables 2.5-2.7.

Our findings suggest that high-ability schools in selective areas raise school aspirations and adult wages, while they lower life satisfaction, compared to mixed-ability schools. Conversely, low-ability schools in selective areas lower school aspirations, while they raise adult wages and self-efficacy. We have attempted to explain this, based on the features of the school system and labour market at the time. Moreover, we find no long-term direct impact of high- or low-ability school attendance on other well-being measures, self-assessed health, risk for cardiovascular disease and risk of chronic stress. As a secondary finding, we note that childhood cognitive and non-cognitive ability measured prior to secondary schooling are significantly associated with later health and well-being. Their role as either direct causal predictor of human capital or mediators between education and human capital should be the subject of further research to explore the determinants of differences among individual outcomes, which would be informative for future educational policy.

Chapter 3

Human capital consequences of missing out on a grammar school education

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

Chiara Pastore*, Andrew M. Jones*†

Abstract

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

Keywords Selective schooling, Human Capital, Health, Fuzzy regression discontinuity design.

JEL I1, I26, I28, C21

*Department of Economics and Related Studies, University of York, United Kingdom.
†Centre for Health Economics, Monash University, Australia.

3.1 Introduction

In 2018, the UK government announced the first £50 million round of a £200 million fund for an expansion in grammar schools, public and selective high-quality institutions. The move lifted a ban on the creation of selective schools that had been in place since 1998. Proponents of tracking policies, that allocate students to different classes or schools on the basis of ability, maintain that they reward talent regardless of socio-economic background. Opponents, on the other hand, raise the concern that entry criteria are skewed in favour of children from affluent backgrounds and children whose ability develops earlier, since they are disproportionately more likely to do well in entry tests.

Being assigned to a school that selects its pupils on the basis of ability is likely to affect long-term outcomes by providing a pool of more able peers and better networks for the future, or through a curriculum with more academic content, which could facilitate later admission to better higher education programmes and the development of a broader set of skills. Moreover, there is evidence that more qualified teachers seek schools with higher ability pupils and that better resources are allocated to these schools (Levačić and Marsh, 2007; Pop-Eleches and Urquiola, 2013). Recognising its relevance to the current policy context, this paper explores the medium- and long-term effect of going to grammar school, compared to its main alternative within a selective system, on a broad range of human capital and health outcomes for individuals of similar prior ability. The analysis is based on data from the National Child Development Study (NCDS), a British cohort study of individuals born in March 1958, who started secondary school in 1969 and whose lives have been followed for 60 years to date.

The literature looking at the effects of selective schooling can be divided into two main strands. A first set of studies compares selective and non-selective systems. Some studies in this category, implementing value-added or instrumental variable approaches to estimate an average treatment effect of selection using NCDS, find a positive effect of grammar school on educational attainment for high ability pupils (Galindo-Rueda and Vignoles, 2005; Harmon and Walker, 2000). These studies have been criticised as unable to deal with selection bias by Manning and Pischke (2006). Other papers have addressed these concerns using administrative and household panel data to compare selective and non-selective areas with matching methods, generally finding no difference in average outcomes, but instead a link between selection and inequality in education and earnings (Atkinson et al., 2006; Burgess et al., 2017, 2019). The second robust set of studies on the effect of selective education has estimated a local average treatment effect for the marginal admitted student, based on regression discontinuity design approaches (Clark, 2010; Del Bono and Clark, 2016; Guyon et al., 2012). Due to data limitations, this second approach has not been implemented before with English country-wide data.

We use a regression discontinuity design (RDD), based on the fact that admission to grammar school was determined by whether a pupil scored above the local education authority (LEA) pass mark in an entry exam, known as 11-plus. As a proxy for the 11-plus score, we use age 11 cognitive tests collected in the NCDS. These closely mirror the three components of the 11-plus and are therefore reliable predictors of grammar school entry. To overcome the issue of limited data on entry score cut-offs for all English LEAs in the 1960s, we proxy pass marks with LEA-specific thresholds estimated directly from the NCDS dataset. Following the structural breaks literature (Bai, 1997), in the same spirit as Card et al. (2008), we select the threshold value that maximises the fit of a model of school assignment, bootstrapping standard errors to account for possible sampling error. This approach identifies the medium-term effect on educational attainment and labour market outcomes, as well as the long-term effect on health and risk of developing illness up to age 50.

This paper contributes to the literature on the long-term effect of grammar schools in England in several ways. We are the first to exploit an RDD strategy to answer this question for all of England. This approach is informative for current policy, since an expansion in grammar school places is likely to affect pupils who were previously on the margin of being admitted. RDD helps us identify the treatment effect for that specific group. Second, we implement RDD in a context with limited information, constructing the assignment variables and score cut-offs from large survey data, in the absence of administrative records. We use several robustness checks to increase confidence in the estimation strategy, confirming previous results obtained with other methodologies. Third, we are able to investigate a broad range of long-term outcomes rather than only educational ones, including long-term health conditions and disease risk. Moreover, the NCDS dataset includes high-quality information on individuals' pre-treatment characteristics, allowing us to analyse background as a competing explanatory variable and to check similarity of individuals near the threshold. Also, since schools are surveyed as part of the study, we can explore how specific school features may explain any effect of school type for long-term outcomes.

For the marginal student, we find a significant and positive effect of grammar attendance on the probability of achieving A-levels, a secondary academic qualification. However, this effect is conditional on having high socio-economic status or high parental interest in education. No effect is found on adult labour market outcomes, health and biomarkers for risk of developing chronic illnesses. This result holds both for a standard RDD approach estimated with local polynomial regressions and when using bias-corrected procedures proposed by Calonico et al. (2017, 2014a). We also find that peer quality could be a significant mechanism to attain better educational qualifications, although a large portion of human capital differences by school type is due to pre-schooling ability and background. Our work draws from historical data and concludes that pupils who narrowly missed out on grammar places in the 1960s missed out on marginally higher chances of better educational attainment, but only if they had a favourable family background. In order to understand the implications of the 2018 Selective School Expansion Fund for current generations of pupils, future evaluations will have to take into account several changes in educational and labour market opportunities faced by younger generations.

The paper is structured as follows. We describe the context and existing literature in Section 3.2, while the data is presented in Section 3.3. Section 3.4 discusses the threshold estimation procedure, accompanied by the necessary conditions for identification and the empirical approach. Section 3.5 presents the results and robustness checks, followed by an exploratory mechanism analysis in Section 3.6. Section 3.7 discusses our findings and Section 3.8 concludes.

3.2 Selective schooling and RDD

Grammar schools have been present in England since the 16th Century, but their formal inclusion in the compulsory state-funded secondary education system dates back to the 1944 Education Act (Jesson, 2013). The Act established a tripartite system of secondary schooling, distinguishing between selective grammar schools, non-selective secondary modern schools and a minority of technical schools. Since then, admission to grammar schools has been determined by performance in the 11-plus test. Until the 1960s, the vast majority of English children would sit the 11-plus at the end of primary school, usually in September of their last year. The exam included language, numerical and reasoning components. It was set locally by Local Education Authorities (LEAs), and the entry mark depended on the number of grammar school places in the area. Rather than setting a given pass score, LEAs usually considered the distribution of scores and then assigned grammar school places to children scoring in the top tail, so the pass mark varied every year. On average, pupils scoring in the top 25% of the distribution in their local area were admitted to grammar school (Bolton, 2017). Area-specific differences in admission included different school capacity constraints, and different policies concerning teacher's recommendations, distance from the school and having other siblings already at the school[1]. Panels of teachers and LEA representatives made decisions on where to allocate students. Those who did not reach the entry score were generally assigned to secondary modern schools, with a less academic focus[2]. Parents could subsequently request a different school or decide to appeal against the panel's decision if they did not agree with it, but this was uncommon. Thus, students with similar ability scores could be assigned to different types of school for two reasons: either because they were close to the cut-off entry score for their area, or because they were from different areas. To the extent that scoring slightly above or below the entry cut-off is random, this paper can isolate the long-term effect of going to grammar school, compared to just missing out on admission, on a broad range of outcomes.

RDD based on assignment test scores has been a popular approach to estimate the causal effect of higher-quality schools on a variety of outcomes in several countries (Abdulkadiroglu et al., 2014; Del Bono and Clark, 2016; Dobbie and Fryer, 2014; Dustmann et al., 2017; Guyon et al., 2012; Kirabo Jackson, 2010; Pop-Eleches and Urquiola, 2013; see Table B1 for more details). Most studies find a positive effect of higher-quality schools on measures of educational attainment. Some however, find no effect (Abdulkadiroglu et al., 2014; Dustmann et al., 2017) or even a negative effect, as in the case of a Swedish reform that made vocational education more academic, which was found to increase dropout rates (Hall, 2012). A highly relevant study by Clark (2010), who implements a regression discontinuity design to investigate the impact of grammar schools on educational outcomes for Yorkshire, a region in England, finds no effect on educational achievement, but a positive effect on university enrolment. In a different study, Guyon et al. (2012) exploit a reform causing an exogenous increase in the number of pupils attending grammar schools in Northern Ireland, and find that average educational outcomes increased following the reform.

Two more RDD studies stand out in the literature because of their focus on long-term outcomes other than educational attainment. Dustmann et al. (2017) study the German secondary school system, exploiting month of birth as a source of quasi-random variation in track assignment at age 10. They find no effect on labour market outcomes such as earnings, probability of employment and occupational choice, and they relate this result to flexibilities in the German school system, which allow for reallocation of students to a more suitable track

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

[2]From 1965 onwards the Circular 10/65 approved by the Labour government encouraged LEAs to move towards a third type, comprehensive schools, catering for students of all abilities. This paper focuses on areas that were still largely selective when the data was collected, and where a large proportion of pupils attending public secondary schools were assigned to either grammar or secondary modern. See Section 3.5.5 for further discussion.

at later grades. A second study, by Del Bono and Clark (2016), is the one most similar to ours, since they investigate long-term effects of elite schools in Scotland, for children born in the 1950s. In addition to educational and labour market outcomes, they study effects on fertility. They find a significant and positive effect on educational outcomes, while they observe higher wages and lower fertility for girls only. Their RDD strategy exploits information on admission exam scores (equivalent to the 11-plus), and entry score cut-offs for the city of Aberdeen.

Given the unavailability of such precise information at Local Education Authority level for England, literature on the effects of grammar schools has been unable to exploit RDD methodologies to answer this question in the English context, except for the study by Clark (2010), which focuses on one region in England and on education outcomes only. The main challenge in the literature has been to deal with selection bias, arising from the inability to control exhaustively for pre-treatment characteristics of treated and control children. Value-added and instrumental variable approaches, such as those in Galindo-Rueda and Vignoles (2005), Harmon and Walker (2000), Jesson (2000), and Kerckhoff (1986), have been criticised by Manning and Pischke (2006) as unable to remove the selection problem, which was evidenced by a placebo-type test, showing a spurious effect of type of school on outcomes prior to school.

More recent studies focusing on the English context have tried to overcome this problem by matching treated and control groups on either individual or local-area characteristics. Among these, Atkinson et al. (2006) find a positive effect of grammar on educational outcomes, while Burgess et al. (2017, 2019) find that selection increases educational and earnings inequality. A different approach in an unpublished paper by Maurin and McNally (2009) looks at the effect of 'comprehensivisation' of the English secondary school system by comparing individuals from the same areas, who experienced two different systems because they belong to two different cohorts. They find a positive effect of grammar education for affected individuals, but also a positive effect of the transition to comprehensive schooling on average educational outcomes. Fewer papers have focused on the health and well-being effects of selective education. Although the raw data shows better adult health for former grammar pupils, the effect of grammar on health is mostly not statistically significant when including all relevant pre-treatment characteristics (Jones et al., 2018, 2012). However, the average could be hiding heterogeneity by ability levels. For instance, the transition to comprehensive schooling has been found to worsen health and smoking for individuals with lower non-cognitive skills only (Basu et al., 2018). We now build further on this evidence by looking at human capital for the group of marginal students affected by the additional places. The effect is policy relevant as it expresses the impact of making it into grammar school versus missing out on the marginal place, for similar levels of initial ability and everything else being equal.

3.3 Data

The NCDS is a longitudinal study of individuals born in the United Kingdom in a single week in March 1958. 98% of all individuals born in England, Scotland and Wales during that week were part of the birth survey, making it nationally representative for that cohort. Following the birth sweep, which contained over 17,000 individuals, surveys were undertaken at ages 7, 11, 16, 23, 33, 42, 45, 50, 55. At the latest sweep the survey still retained over 9,000 individuals (Brown et al., 2016). For the present study, only English and Welsh individuals were included in the sample, since Scotland had a different schooling system in place[3].

3.3.1 Sample and type of school

Information on grammar school attendance is retrieved from the age 16 wave. The sample consists of individuals who went to grammar or secondary modern schools between the ages of 11 and 16, attended by 10% and 20.6% of the NCDS sample respectively. Information on LEA of school, essential for recreating the actual peer group of test-takers from NCDS data, was obtained via special licence access, also from the age 16 wave[4]. Of the total 18,521 individuals for whom there is information in the survey, 5,366 were excluded because they did not have relevant information on school attended, while a further 9,131 individuals were excluded as they did not attend neither grammar nor secondary modern school leaving 4,024 individuals. This large drop in the sample is due to the fact that many areas at the time were already transitioning to a mixed-ability system following the Circular 10/65, which encouraged LEAs to abolish selection by ability at the school level. The sample for threshold estimation thus consists of grammar and secondary modern pupils for whom we also have age 11 cognitive test scores, which yields 3,448 individuals. Due to the inclusion of covariates, and to missing items from surveys at different ages, samples for outcome regressions are always smaller, ranging roughly between 1,450 and 2,800, depending on the outcome[5].

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

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

[5]The reason why we do not use the same sample for threshold estimation and estimation of treatment effect is that a reliable threshold figure needs to capture as much of the actual test-taking population as possible.

3.3.2 Ability

The main variable needed for the identification strategy, age 11 cognitive ability scores, is obtained via principal component analysis (PCA) for the maths, reading and general ability test modules included in the NCDS, following previous literature (Cawley et al., 1997; Galindo-Rueda and Vignoles, 2005; Jones et al., 2011). More details on this procedure can be found in Appendix Section A.3. The cognitive ability index is the first principal component obtained via PCA, which explains approximately 85% of the total variance in the three test scores. This approach is preferred to including the three separate scores for two reasons. First, the test scores are highly related, and including them all would introduce multicollinearity issues in the estimation procedure, increasing standard errors of coefficient estimates (Gujarati, 2004). Second, the three scores obtain similar relative weights for the first principal component obtained via PCA, meaning their contribution to our index is roughly equal. The 11-plus actual total score was also an average of similar language, mathematics and reasoning tests, offering further support to the assumption that our index is a good proxy of 11-plus scores[6]. 97% of NCDS children took the NCDS ability tests between April and July of 1969, thus only a few months after having sat the 11-plus (see Appendix Figure B1). This is usually taken in September of the last year of primary school, which for NCDS children would have been 1968-1969. For our empirical application, it is reassuring that ability tests were still taken in primary school, and not at the start of secondary school, where tests could be affected by the treatment.

3.3.3 Outcomes

Our broad range of outcomes allows us to build a rounded picture of the consequences of school quality for the individual, covering the education, labour market and health domains. Education outcomes are binary variables equal to 1 for having obtained any A-levels (or equivalent) and having a university degree, asked at age 23. Labour market outcomes are measured at age 33 and are all retrieved from survey questions. They include binary variables for being unemployed, receiving state benefits (excluding child benefits) and gross hourly wage, which is calculated from weekly hours worked and weekly wages and then log-transformed for regression analysis. Self-reported health outcomes include age 50 self-assessed health and low malaise scores, which have been validated as good predictors of

[6]Similarity of the two tests is also confirmed by comparing the individual tasks. For the interested reader, we suggest comparing the sweep 2 NCDS ability test, available on the Centre for Longitudinal Studies website at https://cls.ucl.ac.uk/wp-content/uploads/2017/07/NCDS2-Guide-to-the-Dataset.pdf, with sample tests in Bristow (2016), containing a range of tests used from the 1950s onwards.

physical and mental health respectively (Eriksson et al., 2001; Kaplan and Camacho, 1983; Rodgers et al., 1999). Self-assessed health (SAH) is measured on a 5-point scale ranging between 'Excellent' and 'Very poor', and converted to a binary variable equal to 1 if SAH is 'Excellent' of 'Very good', and 0 otherwise. Malaise score is measured via the 9-item Malaise Inventory, developed by Rutter et al. (1970) (see Appendix Section A.4). In regression analysis, we use a binary variable equal to 1 for low malaise (score 0-2 out of 9) and 0 otherwise, increasing in good mental health. We also include the following three biomarkers measured at age 45, all increasing in the risk of cardiovascular disease and health complications (Benzeval et al., 2014). Body mass index (BMI) is calculated as weight in kg divided by squared height in meters, while cholesterol ratio (mmo/L) and triglyceride levels (mmo/L) are retrieved from blood samples.

3.3.4 Background information

The main advantage of cohort studies such as NCDS, compared to administrative data, is that we can control for specific individual characteristics measured early on in childhood. We account for childhood non-cognitive skills in all regressions, given their demonstrated importance for long-term outcomes (Kautz et al., 2014). They are measured at age 11, prior to starting secondary school, by asking primary school teachers questions on the twelve behavioural dimensions that are part of the Bristol Social Adjustment Guide (BSAG)[7]. We then converted the total BSAG score to a variable bounded between 0 and 1 for convenient interpretation. The vector of covariates further includes sex; mother's interest in child education on a 4-point scale and father's socio-economic status on a 5-point scale; whether the mother was smoking during the fourth month of pregnancy; a childhood morbidity index for the cohort member. The morbidity index was constructed following previous literature, by summarising information on twelve categories of childhood conditions up to age 7 into a variable bound between 0, indicating no morbidity, and 1, indicating highest morbidity (Jones et al., 2011; Power and Elliott, 2006).

3.3.5 Attrition over time

Attrition in the data over time can introduce bias in treatment effect estimation when it is correlated to specific individual characteristics. In Appendix Table B2, we compare average characteristics of individuals who dropped out from the sample to individuals for whom we observe at least one outcome. Approximately one in six NCDS children attending grammar and secondary modern, for whom we

[7] More details can be found in Chapter 2 of this thesis and in Shepherd (2013).

also have age 11 ability scores, drops out from the sample. On average, individuals who dropped out have lower childhood cognitive skills, and a somewhat less favourable background in terms of all covariates considered, except for childhood morbidity, which is similar. However, we argue that attrition does not pose a large threat to our results. First, since we estimate the effect of grammar around the cut-off for individuals of similar cognitive ability, individuals who dropped out of the survey are likely to fall outside of the bandwidth used for treatment effect estimation. A valid concern could then be that the location of the cut-off is affected by attrition. As explained in detail in Section 3.4, the location of the cut-off is estimated including individuals for whom we have information on age 11 ability score and age 16 school type and LEA of residence, but who could subsequently drop out. We find that the probability of being observed in the age 16 survey is positively correlated with age 11 cognitive ability scores in the total sample. Since LEA of residence is functional to the definition of the cut-off, and it is retrieved from the age 16 survey, this means that the sample on which we base our threshold estimate has higher average cognitive ability than the population of 11-plus test takers[8]. However, this does not in principle shift the location of the cut-off, since location depends on an existing (although unknown) discontinuity in the probability of attending grammar. Finally, it is particularly reassuring that the probability of dropping out from the sample is smooth around the cut-off, as shown in Figure B2 and Table B4 in the Appendix.

3.4 Methods

3.4.1 Threshold selection

The threshold for each local education authority, necessary for identification, is estimated directly from the data. We do not have information on actual grammar admission marks by LEA, although we can infer them from the ability test scores of pupils attending grammar and secondary modern schools within each LEA in the NCDS. Asymptotic theory for estimated discontinuity points comes from the financial structural breaks literature (Bai, 1997; Bai et al., 1998) and more recently inference procedures have also been developed for a broader range of settings (Porter and Yu, 2015). Reassuringly for the present application, the key finding from this literature is that treatment effect in the RDD can be efficiently estimated in presence of discontinuity points estimated from the data (Porter and Yu, 2015). Thus, for each LEA, we first run probit models for grammar

[8]See Table B3 in the Appendix. We thank an anonymous journal reviewer for suggesting this check.

attendance with a single regressor $1[A_i \geq c]$. This is an indicator function for whether individual i's ability A_i is equal or greater than a threshold c, for a pre-specified range of possible thresholds $c \in [-0.2, 1.5]$. A grid search for the highest log-likelihood achieved by these models for each LEA then yields the chosen LEA-specific threshold c_{LEA}.

The approach is close in spirit to Card et al. (2008), who look for the presence of tipping points (i.e. discontinuities) in changes in the share of white population in US neighbourhoods over time, as a function of share of other minorities. Their hypothesis is that discontinuities in these changes are located at specific values of the base-year minority population share. The study analyses a larger sample than ours, and can therefore estimate the threshold on a subsample of the total available sample, and then use the estimated tipping point on the remaining observations, reducing the risk of sampling bias[9]. Given the smaller sample size in the present case, we do not split the sample, but we bootstrap the threshold search procedure, simulating the estimation sample and subsequent threshold search 500 times, and then use bootstrapped standard errors for statistical inference for the treatment effect estimators. Compared to Card et al. (2008), we have the advantage that an LEA-specific discontinuity is known to be present in the school assignment function (since there was a pass mark for grammar school entry), although its exact location is unobservable to us.

3.4.2 Identification

Identification of treatment effects in the RDD is based on the assumption that pupils scoring near the LEA-specific pass mark, where the discontinuity is observed, have similar baseline characteristics. If this assumption holds, near this threshold treatment assignment is as good as random, and differences in long-term outcomes are caused exclusively by treatment (Lee and Lemieux, 2010). Estimating the extent to which treatment alone causes these differences yields a local average treatment effect (LATE) for the group of compliers. These are the individuals in proximity of the threshold, who are assigned to treatment by virtue of scoring above the cut-off.

Identification is based on two stages. The first stage models school assignment. We denote the treatment variable as $G_i \in \{0, 1\}$, where $G_i = 1$ indicates grammar attendance. Following Lee and Lemieux (2010), treatment assignment, which is assumed to change discontinuously at a LEA-specific cut-off level c_{LEA} of the

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

assignment variable A_i, ability test score, is modelled as:

$$G_i = \gamma + \theta 1[A_i \geq c_{LEA}] + h(A_i - c_{LEA}) + \upsilon_i, \tag{3.1}$$

where $1[A_i \geq c]$ is the indicator variable for equal or greater than the threshold, $h(.)$ is a generic function of individual's distance from the pass mark and υ a random error term. We expect this function to be non-deterministic, since children with the same score may be assigned to different schools. There are at least two reasons why we see this in this context. First, because of imperfect compliance by the school, due to the fact that 11-plus test scores by themselves did not grant access to grammar school, but other factors contributed to admission too, as discussed in Section 3.2. Second, additional fuzziness may be caused by limitations of the data, as noted by Card and Giuliano (2016) in their analysis of US school data. Recall that we are using a proxy of actual 11-plus scores, and children may have performed differently in NCDS tests than in the 11-plus. Additionally, although we provide evidence in support of the threshold selection procedure performed, we acknowledge that not observing actual threshold values may further increase fuzziness in the first stage. In consideration of these characteristics of the treatment assignment function, we rely on a fuzzy RDD, where the probability of treatment assignment does not jump sharply from 0 to 1 at the cut-off value of the ability test score A_i, but by a smaller amount, for the reasons listed.

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

$$Y_i = \alpha + \beta G_i + f(A_i - c_{LEA}) + \epsilon_i. \tag{3.2}$$

where Y_i are human capital outcomes, β the treatment effect of interest, $f(.)$ a function of distance from the threshold and ϵ a random error term. The vector of individual-level covariates X_i has been suppressed for ease of notation, but these are assumed to enter both Equations (3.1) and (3.2). Following previous literature, we propose an RDD estimator analogous to a Wald estimator in 2SLS procedures (Hahn et al., 2001; Lee and Lemieux, 2010), so that average treatment effect is identified by the change in the outcome variable produced by a change in the assignment variable (i.e. the reduced form), divided by the change in the first stage, expressing treatment as a function of the assignment variable.

Interpretation of the Wald estimator as an average treatment effect is conditional upon the following assumptions. The first necessary condition is that the assignment variable cannot be precisely manipulated by the individuals in the sample. In this setting, we use a proxy for 11-plus test scores. Precise manipulation of the assignment variable seems unlikely, since individuals have no incentive to change their NCDS ability score depending on the local area grammar school pass mark[10].

The second necessary condition is that other pre-treatment covariates are smooth functions of the assignment variable, to rule out that the treatment effect estimate is confounded by discontinuities in other variables. A first simple check consists of plotting each covariate against the assignment variable for the whole sample, to exclude any discontinuities. Secondly, following Card and Giuliano (2016), we construct an index of all covariates and plot it against the assignment variable, grouping observations into fixed-size bins. The index consists of the fitted values obtained from regressing each outcome on all pre-treatment covariates. Any jumps in the plotted index would invalidate unbiasedness of the treatment effect estimate. Neither the simple covariate plots nor the index graphs contradict the covariate smoothness assumption, further supporting the idea that the discontinuity in treatment assignment is only caused by the assignment variable (see Figures B3 and B4 in the Appendix).

Under monotonicity of the instrument (i.e. $\theta > 0$ for all i, or $\theta < 0$ for all i) and the other verified assumptions (no precise manipulation and smoothness of covariates as a function of the assignment variable), β in Equation (3.2) can formally be interpreted as a local average treatment effect (LATE) for individuals in the proximity of the threshold (Lee and Lemieux, 2010). That is, $\beta = E\{Y_i(1) - Y_i(0)|A_i = c\}$, where $Y_i(1)$ and $Y_i(0)$ denote treated and untreated outcomes respectively. As specified above, the LATE is calculated for compliers, who are those individuals who attend grammar rather than secondary modern because their score allows them to be just above the cut-off for their LEA. Estimating the long-term effect of grammar attendance for this group is interesting because it allows us to understand the effect of funding an expansion in grammar school places, as students on the margin are those who are most likely affected by it. At the same time, under the stated assumptions, the estimation strategy allows us to isolate the effect of grammar from pre-schooling ability and other pre-treatment confounders.

[10] In standard RDD, a condition for the no precise manipulation assumption is that the density of the assignment variable should be reasonably smooth around the threshold, routinely tested via McCrary tests for the assignment variable (McCrary, 2008). The test looks for discontinuities in the density function of the assignment variable, the absence of which supports the smooth density assumption. However, in the present case, the test may be inadequate, since the LEA-specific threshold is estimated based on goodness of fit measured on the available data, and we would expect this to be reflected in the density function of the constructed assignment variable. Instead, we check whether our main results are sensitive to excluding a portion of selected observations around the threshold, an approach known as donut-RDD (Barreca et al., 2016).

3.4.3 Implementation

Choice of bandwidth, kernel and polynomial

A way to ensure similarity between treated and control group is to accurately choose the neighbourhood around the cut-off, from which observations for the estimation of β are drawn. Following the RDD literature, we refer to this neighbourhood as the bandwidth h (Lee and Lemieux, 2010). The smaller the bandwidth, the higher the number of observations excluded, and the higher the probability that the similarity assumption holds for individuals whose assignment variable lies within $[c-h, c+h]$. Bandwidth selection then incurs a trade-off between precision and bias, since larger windows around the cut-off will yield estimates with lower variance but potentially higher bias (Lee and Lemieux, 2010)[11]. A popular approach is to choose the bandwidth that minimises an approximation of the mean squared error (MSE) of the local linear estimator of β, $MSE = (\hat{\beta} - \beta)^2$ (Calonico et al., 2014a; Imbens and Kalyanaraman, 2012).

While an MSE-optimal bandwidth is generally recommended for its point estimator performance, a recent body of literature has shown it is inadequate for inference procedures (Calonico et al., 2018a, 2017, 2018b, 2014a; Cattaneo and Vazquez-Bare, 2016). The argument in a nutshell is that in using the estimated MSE-optimal bandwidth for neighbourhood selection (i.e. $[c - h_{MSE}, c + h_{MSE}]$) we are introducing a misspecification bias, but this bias makes inference based on observations within the neighbourhood and the resulting point estimator invalid (Cattaneo and Vazquez-Bare, 2016). Since the MSE-optimal bandwidth is usually too large for inference, one possible approach would be to simply shrink it, a procedure known as undersmoothing (Cattaneo and Vazquez-Bare, 2016; McCrary, 2008). Instead, Calonico et al. (2014a) propose a robust bias-correction procedure (CCT correction from here onwards) for bandwidth selection, which estimates bias and then adjusts both the regression discontinuity (RD) point estimate and variance estimator. The CCT correction allows for robust confidence intervals, less sensitive to small bandwidth variations and accounting for the variability introduced when correcting for the estimated bias term in the treatment

[11] Bias arises because the further from the threshold, the larger the differences between individuals, not only due to treatment but also due to other confounders.

effect estimator[12]. Alongside our main specification, with local regressions implemented using the MSE-optimal bandwidth selection approach, we thus propose an alternative with CCT bias correction, in order to investigate sensitivity of the analysis to this procedure[13]. For a discussion of other possible criteria for bandwidth selection, including cross-validation, Fan and Gijbels' (1996) bandwidth selector and local randomization neighbourhood selection methods, see Cattaneo and Vazquez-Bare (2016).

The empirical implementation of RDD further requires kernel functions and polynomials to be chosen. The kernel function assigns weights to observations depending on their distance from the threshold, in order to provide optimal treatment effect estimates. While a triangular kernel, assigning highest weight to observations near the threshold, is intuitively appealing, both Lee and Lemieux (2010) and Card and Giuliano (2016) find no important efficiency losses from using uniform kernels. Since uniformity also makes for simpler computation and interpretation of results, we rely on bandwidth selection to select similar observations, and use a uniform kernel for the main strategy, thus giving equal weight to all observations. A second choice is order of polynomial to be used, given that introducing higher order terms for distance from the cut-off often improves fit of the first stage regression. However, since recent literature has shown that regression discontinuity analysis based on high-order polynomials may be misleading (Gelman and Imbens, 2019), we use a 1st-order polynomial, and only introduce interaction terms between threshold and distance from threshold in the main specification. This accounts for the intuition that not only the intercept but also the slope of the average of the dependent variable as a function of the assignment variable may be different above the threshold[14].

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

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

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

Empirical specification

The empirical counterparts to Equations (3.1) and (3.2) are then as follows:

$$G_i = \gamma + \theta_0 T_i + \theta_1 D_i + \theta_2 D_i \times T_i + X_i' \eta + \upsilon_i \tag{3.3}$$

$$Y_i = \alpha + \beta_0 G_i + \beta_1 D_i + \beta_2 D_i \times T_i + X_i' \xi + \epsilon_i, \tag{3.4}$$

where $T_i = 1[A_i \geq c_{LEA}]$ is an indicator for above the threshold, $D_i = (A_i - c_{LEA})$ indicates the distance between the individual's ability test score in the NCDS and the LEA-specific threshold, $D_i \times T_i$ is the interaction between distance and the discontinuity and X_i is a vector of individual characteristics. We estimate Equations (3.3) and (3.4) as the first and second stage of a two-stage least squares regression, with G_i as the endogenous treatment variable, bootstrapped standard errors clustered at the LEA level, and only on the sample selected by using the MSE-optimal bandwidth. We present results without bias correction first, and then with bias correction as a robustness check, following the procedures proposed by Calonico et al. (2014b).

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

3.5 Results

3.5.1 Summary statistics

Summary statistics for baseline characteristics and outcomes by type of school for the whole sample are shown in Table 3.1. Grammar pupils display higher average cognitive and non-cognitive skills, a higher proportion of female pupils and a more advantaged parental background than secondary modern pupils, while no difference is shown in the childhood morbidity of the two groups. Grammar pupils also display better outcomes in adulthood across all domains considered. They have a higher chance of getting A-levels (50% vs 3%) and a university degree (31% vs 2%); they are half as likely to be unemployed or have received state benefits at 33 (excluding child benefits), and they have a higher hourly wage. The long-term

Table 3.1: Descriptive statistics by secondary school attended

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

Mother interest in child education is on a scale from 1-Little interest to 4-Over concerned. Father's SES is on a scale from 1-Low to 5-High. Maternal smoking during pregnancy is on a scale from 1-Non-smoker to 4-Heavy smoker. Healthy ranges for the biomarkers are: <25 for BMI, <5 for cholesterol ratio, <1.7 for triglycerides (Fuggle, 2018).

health of grammar pupils is also better. At age 50 they are more likely to display high levels of self-assessed health (SAH) and low levels of malaise. Their risk of cardiovascular disease and comorbidities at age 45 is also lower, as shown by the lower average levels of BMI, cholesterol ratio and triglycerides.

As a preliminary analysis, we conduct OLS regressions for each outcome with the treatment indicator (grammar) as a single regressor first, and then with additional covariates, estimated on the whole sample of pupils. Table 3.2 shows that the coefficient for grammar is highly significant for all outcomes, except for the low malaise dummy. However, in line with expectations, some of this association is accounted for by adding covariates to the model. The association is explained mainly by cognitive skills, but also sex, mother's interest in child education and father's socio-economic status (results available on request). The LATE estimate we present in the next section aims at isolating the effect of grammar for the group of individuals who are close to the threshold. Focusing on this neighbourhood helps netting out the effect of cognitive skills and other factors, since in proxim-

Table 3.2: All outcomes: OLS regressions for whole sample.

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

Standard errors clustered at LEA level in parentheses. Covariates included and omitted from the table.
* $p < 0.05$, *** $p < 0.001$

ity of the threshold these do not vary discontinuously, as shown in Section 3.4.2, while the probability of treatment does. Table B5 in the Appendix displays mean and standard deviation of included covariates by type of school, contrasting the whole sample to the complier group, defined as the sample of observations within MSE-optimal bandwidths (we illustrate the sample used for A-levels regressions only). As expected, all covariates are more similar among compliers than in the whole sample.

3.5.2 Discontinuity in grammar school assignment

Figure 3.1 displays LEA-specific threshold values selected by the threshold search procedure outlined in Section 3.4.1. Higher estimated thresholds are not necessarily found in areas with the least grammar places. Instead, the location of the threshold reflects the observed LEA-specific discontinuity, as inferred from the average ability of grammar pupils from each LEA in our sample. Figure 3.2 plots grammar attendance in pre-specified bins, for different levels of the distance between the assignment variable and the LEA-specific cut-off, roughly expressing the probability of grammar attendance as a function of the distance variable. The figure shows a jump in the probability of treatment when the assignment variable is near the threshold, indicated by the dashed vertical line, where $A_i - c_{LEA} = 0$. After this threshold, the average probability of treatment increases by approximately 0.4-0.5, and it keeps rising up to 1 for the most able individuals. The instrument of interest, the indicator variable $1[A \geq c_{LEA}]$, is highly predictive of the first stage and sizeable. Estimating the first stage using each of the 10 outcomes in turn, including all covariates, we find a coefficient on the instrument pointing to an approximate 0.45 increase in the probability of attending grammar at the discontinuity, as already shown by graphical evidence for the whole sample (full results in Table B6 in the Appendix). The distance variable is only significant when interacted with the threshold indicator, and only for the education and labour market outcomes and for the malaise indicator. This makes sense, since below the threshold the probability of attending grammar is close to zero, regardless of the distance. The probability of attending grammar is also positively and significantly associated with non-cognitive skills, mother's interest and father's SES for most of the samples considered. Yet, as long as these covariates are smooth around the threshold, this association does not threaten identification of the treatment effect of interest.

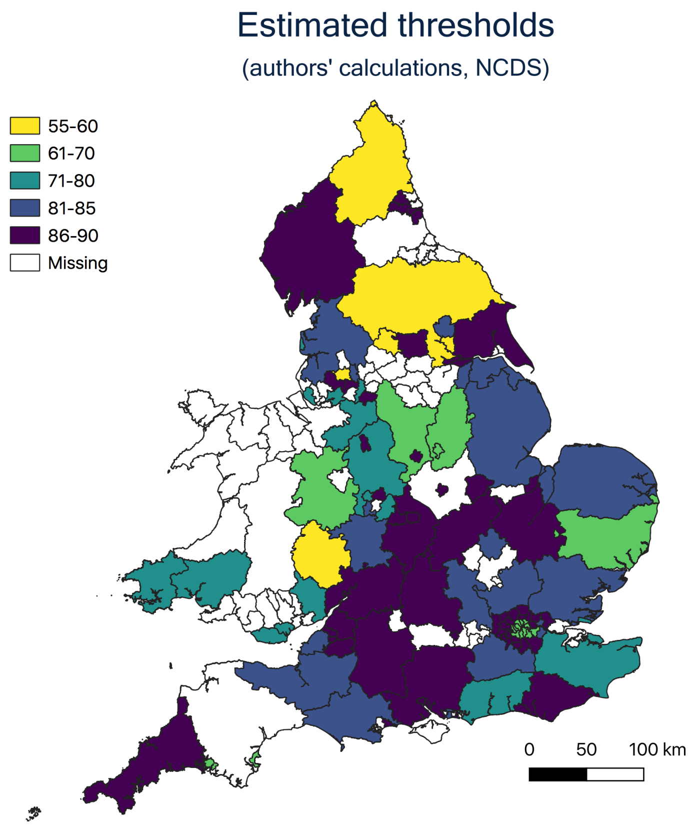


Figure 3.1: Threshold values estimated through our "structural breaks" approach, for the English and Welsh LEAs in the data. The values are expressed in terms of the percentile of the NCDS cognitive ability distribution that would give access to a grammar school place in 1969.

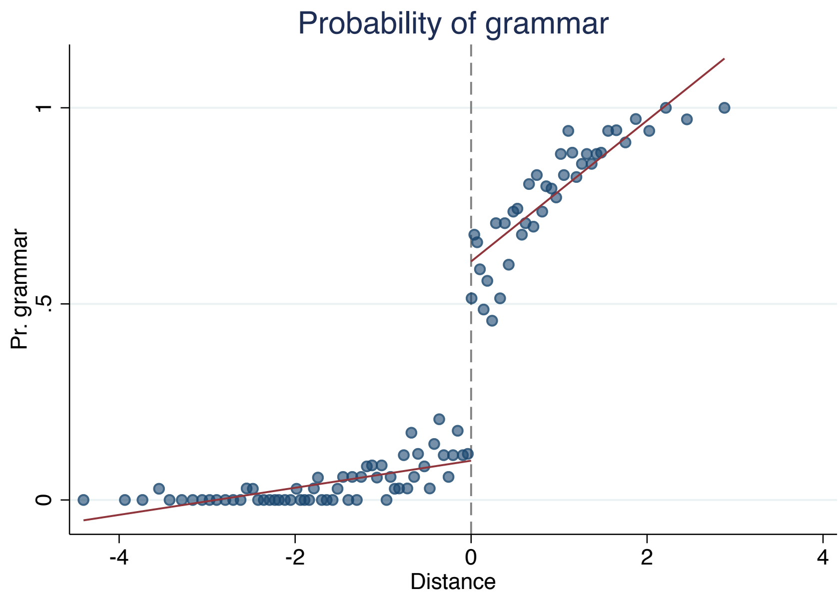


Figure 3.2: Scatter graph for probability of grammar school attendance as a function of the distance D_i between the assignment variable A_i and the LEA-specific threshold c_{LEA}. Observations are grouped in 50 bins, yielding average probability of treatment within the group. Average bin size is 69 (N=3448).

3.5.3 Effect of grammar schools for the marginal student

Figure 3.3 shows scatter graphs for each outcome, so that each dot represents the average outcome for groups of individuals at similar levels of the distance variable. While all outcomes display a relationship with the assignment variable in the expected direction, none shows a sharp jump at the threshold value of zero. The probability of achieving A-levels and a degree display a fairly similar pattern, increasing steeply and linearly after the threshold[15]. Most other outcomes only show an average improvement after the threshold, more evident for cholesterol ratio, triglycerides and the indicator of excellent or very good SAH, but no evident discontinuity. In line with this, Appendix Tables B7 and B8, displaying reduced-form regressions for all outcomes, show that the discontinuity in treatment assignment is only significantly associated with educational outcomes.

Tables 3.3 and 3.4 show LATE estimates for education, labour market and health outcomes, obtained via estimation of Equations (3.3) and (3.4), estimated for the sample within MSE-optimal bandwidths, without CCT bias correction. Standard errors are bootstrapped to account for potential noise in the estimation of the threshold and clustered at LEA level to account for possible heteroskedasticity across LEAs. All models include all covariates described in Section 3.3. The difference in total available sample across outcomes depends on missing information for certain outcomes, and on the fact that MSE-optimal bandwidths vary for each outcome (Lee and Lemieux, 2010). Grammar attendance alone increases the probability of achieving A-levels by 26 percentage points ($p < 0.05$) in the group of compliers. The effect of grammar is significant and positive for obtaining a university degree only with conventional standard errors (not shown, $p < 0.05$), while the larger bootstrapped standard errors return a smaller t-test result, indicating a statistically insignificant effect ($t - test = 1.22$, $p = 1.776$). For both A-levels and obtaining a university degree, the interaction variable between distance and threshold indicator is positive and significant. This indicates a positive change in the slope of the probability of both outcomes as a function of the ability score after the threshold, already shown in Figure 3.3. This result is consistent with the fact that grammar schools were the track with highest academic focus.

The estimated LATE of grammar school for adult labour market, health and disease risk outcomes is not significant, despite being mostly of the expected sign. The sign of grammar is negative for the probability of receiving benefits, cholesterol ratio and triglycerides, while it is positive for hourly wages and self-

[15]Something similar is observed by Dong (2018) in respect to the data used by Del Bono and Clark (2016), who also estimate the impact of elite schooling, but in Scotland and on different outcomes.

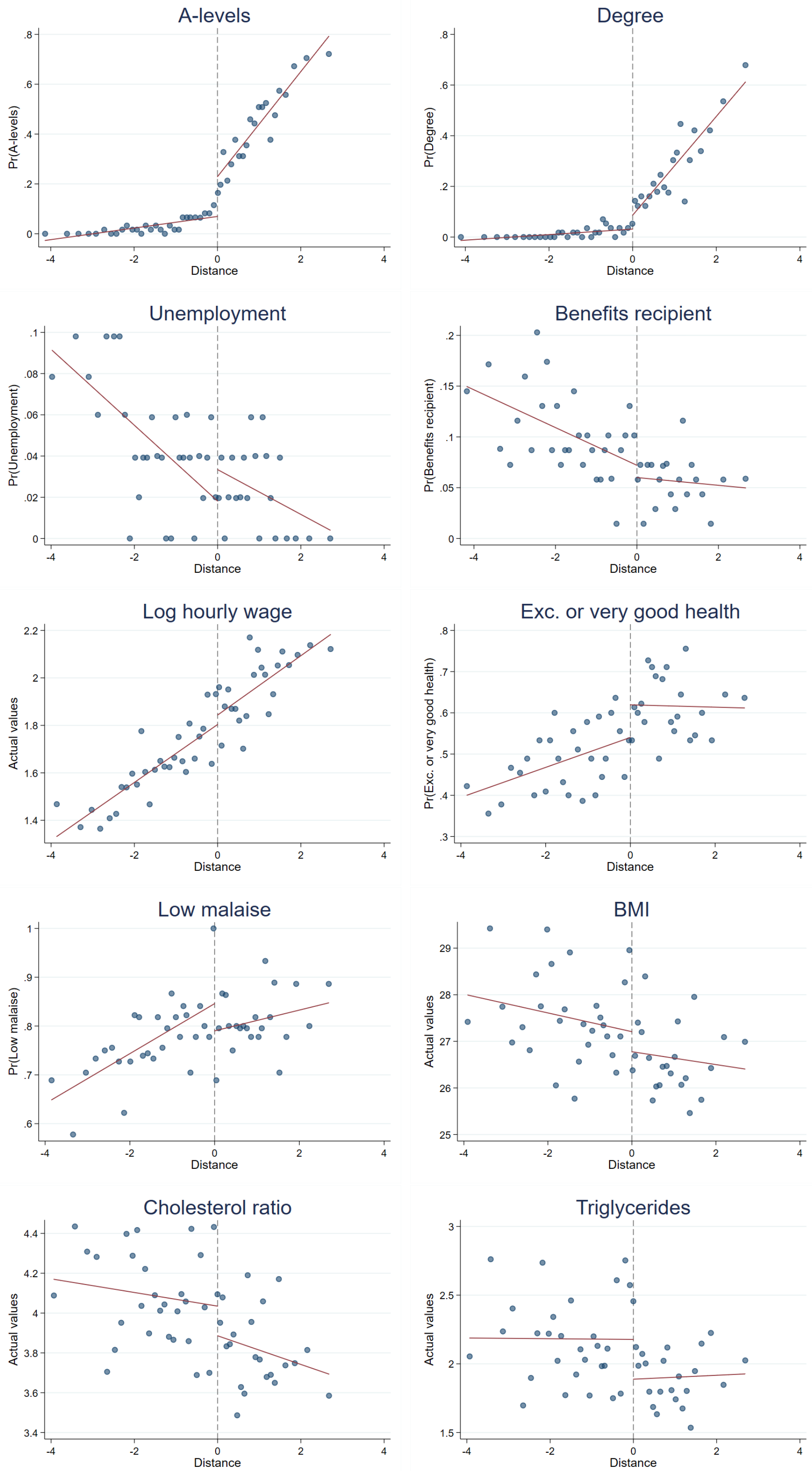


Figure 3.3: Outcomes as a function of distance between the assignment variable and the LEA-specific threshold, $D_i = A_i - c_{LEA}$. Observations are grouped in 50 bins, yielding average outcome within the group.

Table 3.3: Human capital outcomes: local polynomial regressions with pre-selected bandwidth.

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

Bootstrapped standard errors clustered at LEA level in parentheses.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

assessed health. The grammar coefficient for unemployment is close to 0, while the one associated to the probability of scoring low malaise is negative. The interaction between distance and the threshold indicator is significant for BMI and triglycerides level. As anticipated by the OLS regression tables, other background covariates play a prominent role in determining a variety of human capital outcomes. Non-cognitive skills, sex, mother's interest in child education, father's SES and child morbidity index are all significantly associated with some of the outcomes. Selection effects appear to originate early in utero, supporting the fetal origins literature (Almond and Currie, 2011): higher frequency of smoking during pregnancy by the mother shows a positive and significant association with unemployment in adulthood and triglycerides levels. However, smoking during pregnancy is also significantly associated with lower probability of being a benefits recipient, lower malaise and lower cholesterol ratio levels. Finally, higher

Table 3.4: Health outcomes: local polynomial regressions with pre-selected bandwidth.

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

Bootstrapped standard errors clustered at LEA level in parentheses.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

childhood morbidity is associated with lower probability of obtaining A-levels and a degree, higher probability of unemployment and lower hourly wages at 33, and higher BMI and triglycerides.

Table 3.5 displays results of the fuzzy RDD estimation, implementing the CCT bias-correction procedure[16]. Results are fairly consistent with those estimated without bias correction. The CCT bias-corrected MSE-optimal bandwidths are generally smaller, except for hourly wage, cholesterol ratio and triglyceride levels. With the bias-correction procedure, grammar attendance increases the probability of obtaining A-levels by 25 percentage points ($p < 0.01$). Again, the coefficient for university degree is significant with non-bootstrapped standard errors ($p < 0.05$), while it becomes insignificant with the bootstrapping procedure. As before, the grammar coefficient is not significant for health outcomes and risk of illness. Discontinuity estimates in the grammar assignment function (first-stage estimates), at the bottom of each panel, are all highly significant with bias correction also, offering support to the validity of the study design.

3.5.4 Heterogeneous effects

To explore heterogeneity in the effect of grammar school, we additionally provide results for the specification with treatment interacted with sex, father's SES and mother's interest in child education, in Tables 3.6 and 3.7. In panel A of Table 3.6, the grammar coefficient on A-levels is larger than with the pooled sample, and the coefficient on the interaction $Grammar \times Female$ is negative and significant. Thus, for boys, the base category, attending grammar increases the probability of achieving any A-levels by 30 percentage points, while the effect is 11 points lower for girls. Additional heterogeneity analysis is carried out by background characteristics in panels B and C of Tables 3.6 and 3.7. Grammar attendance only has a positive effect on probability of achieving A-levels for children whose father has high SES (18 percentage points increase) and whose mother's interest in child education is high (12 percentage points increase). Conversely, the effect of grammar attendance for the base categories, children with low SES or low mother interest, is not significant, indicating that grammar school attendance only gives an educational advantage to advantaged groups. Again, results for health do not show a significant impact of grammar school, except for cholesterol ratio, which is significantly and negatively impacted for girls and positively for children with high mother interest.

To explore heterogeneity in access by individual characteristics, we additionally plot the probability of grammar attendance against ability scores separately

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

Table 3.5: All outcomes: bias-corrected model with pre-selected bandwidth.

	(1)	(2)	(3)	(4)	(5)
	A-levels	Degree	Unemployed	On benefits	Log gross income
Grammar	0.2458**	0.1773	-0.0245	-0.0644	0.1317
	(0.0948)	(0.1542)	(0.0663)	(0.1144)	(3.2839)
Robust 95% CI	[.038 , .401]	[.04 , .411]	[-.147 , .096]	[-.223 , .024]	[-.192 , .502]
Bandwidth	1.3036	0.6716	0.8144	1.5038	1.4889
Left of c	622	315	357	818	454
Right of c	741	397	417	916	533
Available obs.	2505	2352	2116	2816	1553
First-stage estimate	0.4579***	0.4592***	0.4640***	0.4492***	0.4586***
First-stage conv. s.e.	(0.0405)	(0.0551)	(0.0523)	(0.0351)	(0.0464)
	(6)	(7)	(8)	(9)	(10)
	SAH	Malaise	BMI	Chol	Trig
Grammar	0.0574	-0.2175	-1.0928	-0.4546	-0.5793
	(0.2407)	(0.1584)	(2.4020)	(0.5542)	(0.9435)
Robust 95% CI	[-.312 , .374]	[-.519 , -.036]	[-4.633 , 2.121]	[-1.231 , .515]	[-1.753 , .901]
Bandwidth	0.7865	0.6958	0.7834	1.1720	1.2202
Left of c	301	261	292	360	379
Right of c	380	339	370	430	452
Available obs.	1869	1865	1786	1498	1500
First-stage estimate	0.4936***	0.5218***	0.4835***	0.4240***	0.4320***
First-stage conv. s.e.	(0.0563)	(0.0592)	(0.0553)	(0.0509)	(0.0496)

Bootstrapped standard errors clustered at LEA level in parentheses. Covariates included and omitted from the table.
** $p < 0.01$, *** $p < 0.001$,

for the following subsamples: girls and boys, high and low father SES, high and low mother interest in child education. Patterns in Appendix Figure B5 are similar to the first stage for the pooled sample, except for a larger jump at the threshold for high-SES and high mother interest children. This denotes a higher probability for these subgroups of being admitted to grammar when scoring at or just above the threshold, which is consistent with literature showing that pupils from a more deprived background are under-represented in grammar schools, even when taking prior ability into account (Andrews et al., 2016; Burgess et al., 2018; Cribb et al., 2013).

3.5.5 Robustness checks

We include a battery of robustness checks of our approach. We first implement local polynomial regressions for a placebo cut-off, in order to check that the first stage is not predictive of grammar attendance at other points of the distribution of the assignment variable. We run all outcome regressions at a fictional cut-off of 0.2[17]. Table 3.8 shows that treatment effect estimate is not significant for any outcomes. Moreover first-stage estimates are not significant either, meaning no discontinuities are detected at 0.2 in the distance variable. Figure 3.4, plotting the probability of grammar against ability scores with different placebo thresholds, also shows that the discontinuity in treatment assignment is not detected once we move away from the threshold used in the main specification.

As a second robustness check, we re-estimate the model while excluding a set of observations around the threshold, a procedure known as donut exclusion or donut-RDD. The density of the distance from LEA-specific threshold $D_i = A_i - c_{LEA}$, in Figure 3.5, presents a concentration of observations around the threshold, and we attempt to show that our results are not sensitive to excluding these. MSE-optimal bandwidth selection is operated again after excluding the observations within 0.8 from the threshold, dropping 4% of observations (n=148)[18]. Both the specification with and without CCT bias correction confirm the main results: A-levels and obtaining a degree are positively and significantly associated with grammar attendance, and coefficients are slightly larger in magnitude, which we would expect, since individuals become less similar as we move away from the threshold (see Table 3.9 in main text and Table B10 in the Appendix). Confirming previous results, grammar is not significantly associated with labour

[17] Placebo tests executed with CCT bias correction are presented in Appendix Table B9

[18] The portion of observations excluded for the donut RDD check is chosen based on excluding spikes from the density plot. Small variations in the number of observations excluded did not produce dissimilar results.

Table 3.6: Human capital outcomes: local polynomial regressions with pre-selected bandwidth and treatment interacted with individual characteristics.

	(1) A-levels	(2) Degree	(3) Unemployed	(4) On benefits	(5) Log wage
Panel A: by sex					
Grammar	0.319**	0.157	-0.00557	-0.0541	0.0670
	(0.125)	(0.132)	(0.0585)	(0.100)	(0.230)
Female	0.0191	-0.0172	-0.00722	0.0404*	-0.473***
	-0.0208	-0.0195	-0.0113	-0.0203	-0.0491
Grammar × Female	-0.113**	-0.0220	0.0155	0.0169	0.00463
	(0.0469)	(0.0420)	(0.0248)	(0.0380)	(0.0917)
Distance	0.0212	-0.0123	0.00406	0.00851	0.0860
	(0.0522)	(0.0612)	(0.0300)	(0.0622)	(0.113)
Distance $\times 1[A_i \geq c_{LEA}]$	0.128*	0.138***	0.00138	0.00949	-0.0122
	(0.0574)	(0.0488)	(0.0322)	(0.0465)	(0.0962)
First-stage F statistic	90.414	79.348	56.148	94.635	36.382
Observations	2505	2352	2116	2816	1553
Panel B: by father's SES					
Grammar	0.192	0.123	-0.00609	-0.0419	0.0621
	(-0.132)	(-0.129)	(-0.057)	(-0.0871)	(-0.258)
High father's SES	0.0368**	0.0409***	-0.00302	-0.00695	0.0821**
	(0.0103)	(0.0100)	(0.00617)	(0.00785)	(0.0185)
Grammar × High father's SES	0.188***	0.0701	0.0229	-0.00972	0.0183
	(-0.0439)	(-0.0444)	(-0.0212)	(-0.0293)	(-0.0974)
Distance	0.0284	-0.01	0.00635	0.00852	0.0873
	(-0.0574)	(-0.0612)	(-0.0298)	(-0.0415)	(-0.126)
Distance $\times 1[A_i \geq c_{LEA}]$	0.129+	0.137***	-0.0011	0.00873	-0.0132
	(-0.0701)	(-0.0515)	(-0.0326)	(-0.0373)	(-0.0998)
First-stage F statistic	83.366	75.733	47.674	87.033	29.632
Observations	2505	2352	2116	2816	1553
Panel C: by mother interest					
Grammar	0.188	0.137	0.0245	-0.0452	0.14
	(-0.16)	(-0.136)	(-0.0664)	(-0.107)	(-0.306)
High mother interest	0.00216	0.00827	-0.00282	-0.0146	0.0584*
	(0.0114)	(0.0104)	(0.00679)	(0.00880)	(0.0289)
Grammar × High mother int.	0.118**	0.0303	-0.023	-0.0217	-0.0567
	(-0.0474)	(-0.0468)	(-0.0272)	(-0.0394)	(-0.113)
Distance	0.024	-0.0115	0.000901	0.014	0.0548
	(-0.0708)	(-0.0574)	(-0.0298)	(-0.0538)	(-0.155)
Distance $\times 1[A_i \geq c_{LEA}]$	0.123+	0.132*	0.00542	0.0104	0.0143
	(-0.0629)	(-0.0602)	(-0.0328)	(-0.0492)	(-0.122)
First-stage F statistic	86.540	71.907	57.355	92.487	34.788
Observations	2505	2352	2116	2816	1553

Bootstrapped standard errors clustered at LEA level in parentheses. High father SES corresponds to mid-high or highest. High mother interest corresponds to very interested or over-concerned. Covariates included and omitted from the table. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.7: Health outcomes: local polynomial regressions with pre-selected bandwidth and treatment interacted with individual characteristics.

	(6) High SAH	(7) Low malaise	(8) BMI	(9) Chol	(10) Trig
Panel A: by sex					
Grammar	0.0724	-0.0320	0.0207	-0.180	-0.731
	(0.199)	(0.0999)	(1.957)	(0.446)	(0.780)
Female	0.0592	-0.0798**	-0.393	-0.817***	-0.871***
	(0.0497)	(0.0311)	(0.408)	(0.0801)	(0.0907)
Grammar × Female	-0.0488	-0.0807	-1.079	-0.391**	-0.132
	(0.0976)	(0.0595)	(0.737)	(0.164)	(0.231)
Distance	0.0571	0.0374	1.496	0.213	0.744*
	(0.111)	(0.0666)	(1.121)	(0.233)	(0.378)
Distance $\times 1[A_i \geq c_{LEA}]$	-0.00650	-0.0582	-2.749**	-0.371	-0.883***
	(0.133)	(0.0756)	(1.129)	(0.217)	(0.304)
First-stage F statistic	31.708	45.014	29.475	21.954	20.701
Observations	1869	1865	1786	1498	1500
Panel B: by father's SES					
Grammar	0.000674	-0.0854	-0.722	-0.385	-0.843
	(-0.236)	(-0.136)	(-1.668)	(-0.459)	(-0.616)
High father's SES	0.0440***	0.00859	-0.977***	-0.0658+	-0.121***
	(0.0127)	(0.0143)	(0.139)	(0.0343)	(0.0364)
Grammar × High father's SES	0.101	0.0242	0.383	-0.035	0.11
	(-0.0733)	(-0.0503)	(-0.622)	(-0.138)	(-0.189
Distance	0.0647	0.0341	1.507	0.199	0.751**
	(-0.124)	(-0.0739)	(-0.921)	(-0.21)	(-0.314)
Distance $\times 1[A_i \geq c_{LEA}]$	-0.00565	-0.0493	-2.714**	-0.332	-0.886***
	(-0.109)	(-0.0727)	(-1.069)	(-0.235)	(-0.29)
First-stage F statistic	24.443	35.498	23.238	19.823	18.979
Observations	1869	1865	1786	1498	1500
Panel C: by mother interest					
Grammar	0.111	-0.101	-1.2	-0.829+	-1.118
	(-0.196)	(-0.166)	(-1.822)	(-0.502)	(-0.754)
High mother interest	0.0223	-0.0169	-0.652***	-0.204***	0.206*
	(0.0196)	(0.0149)	(0.245)	(0.0466)	(0.105)
Grammar × High mother int.	-0.115	0.0672	0.822	0.743***	0.222
	(-0.0716)	(-0.0668)	(-0.695)	(-0.21)	(-0.238)
Distance	0.0527	0.0227	1.738	0.286	0.929*
	(-0.0965)	(-0.0793)	(-0.954)	(-0.22)	(-0.412)
Distance $\times 1[A_i \geq c_{LEA}]$	0.0208	-0.033	-3.091**	-0.526*	-1.064***
	(-0.121)	(-0.0754)	(-1.207)	(-0.233)	(-0.332)
First-stage F statistic	35.311	49.653	28.486	21.657	20.899
Observations	1869	1865	1786	1498	1500

Bootstrapped standard errors clustered at LEA level in parentheses. High father SES corresponds to mid-high or highest. High mother interest corresponds to very interested or overconcerned. Covariates included and omitted from the table. $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

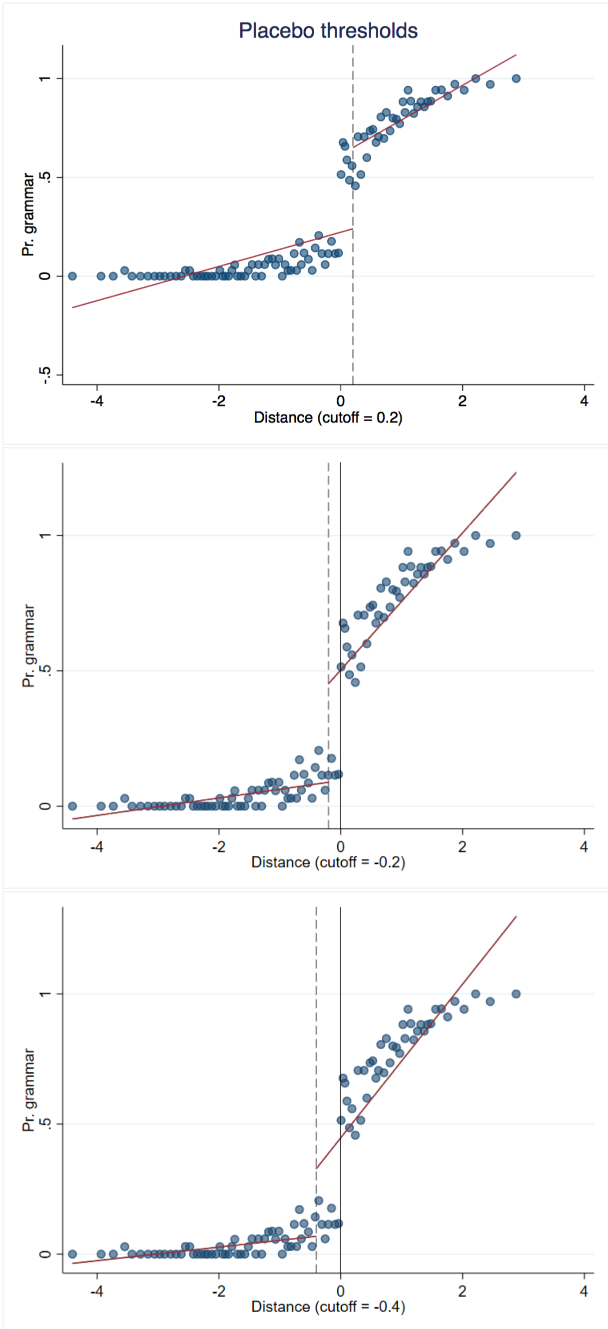


Figure 3.4: Probability of grammar school attendance as a function of distance from the LEA-specific threshold, $D_i = A_i - c_{LEA}$, with placebo cut-offs at 0.2, -0.2 and -0.4, and 50 bins. The scatter points show a sharp discontinuity at 0 and that any other cut-off is unable to account for that.

Table 3.8: All outcomes: local polynomial regressions with placebo threshold at 0.2 and pre-selected bandwidth.

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

Bootstrapped standard errors clustered at LEA level in parentheses. Covariates included and omitted from the table.

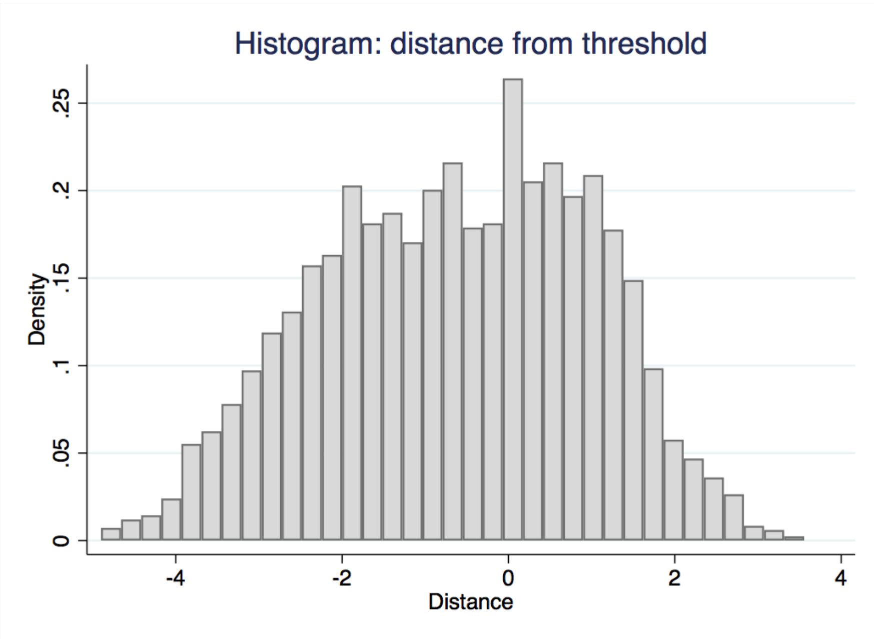


Figure 3.5: Histogram illustrating density of distance between the assignment variable and the LEA-specific threshold, $D_i = A_i - c_{LEA}$.

Table 3.9: All outcomes: local polynomial regressions with donut exclusion around the threshold and pre-selected bandwidth.

	(1) A-levels	(2) Degree	(3) Unemployed	(4) On benefits	(5) Log wage	(6) High SAH	(7) Low malaise	(8) BMI	(9) Chol	(10) Trig
Grammar	0.3486**	0.2198*	0.0354	-0.1968*	0.2118	0.2356	0.1652	-3.2119	-0.3717	-0.2595
	(0.1126)	(0.0933)	(0.0542)	(0.0778)	(0.1718)	(0.1621)	(0.1285)	(1.7917)	(0.3558)	(0.4024)
Distance	0.0162	-0.0448	0.0078	0.1776**	-0.0306	0.0655	-0.0847	3.4039*	-0.0045	0.0796
	(0.0792)	(0.0595)	(0.0378)	(0.0591)	(0.0906)	(0.1103)	(0.0730)	(1.6050)	(0.2405)	(0.2718)
Distance $\times 1[A_i \geq c_{LEA}]$	0.0029	0.1036	-0.0104	-0.1215	0.0277	-0.2522	0.0310	-3.4355	0.0958	-0.0759
	(0.0916)	(0.0707)	(0.0465)	(0.0706)	(0.1147)	(0.1351)	(0.0876)	(1.9603)	(0.2997)	(0.3397)
$1[A_i \geq c_{LEA}]$	0.4710***	0.4671***	0.4755***	0.4598***	0.4861***	0.4891***	0.4557***	0.5016***	0.5112***	0.5111***
	(0.0538)	(0.0522)	(0.0586)	(0.0567)	(0.0545)	(0.0623)	(0.0539)	(0.0746)	(0.0641)	(0.0642)
First-stage F	76.630	79.923	65.741	65.671	79.596	61.713	71.535	45.175	63.510	63.344
Tot. obs. available	905.000	941.000	774.000	881.000	774.000	730.000	850.000	536.000	599.000	598.000
Obs. in bandwidth	2393.000	2246.000	2020.000	2691.000	1481.000	1778.000	1774.000	1696.000	1421.000	1422.000
Bandwidth	0.932	1.006	0.902	0.810	1.216	0.948	1.086	0.733	0.992	0.985

Bootstrapped standard errors clustered at LEA level in parentheses. Covariates included and omitted from the table.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

market and health outcomes[19].

Thirdly, we show that our approach successfully isolates the effect of grammar school attendance from that of background, by estimating the effect of the discontinuity in grammar attendance on placebo outcomes prior to secondary school. This check is in the same spirit as an influential test conducted by Manning and Pischke (2006), showing that a set of empirical approaches did not deal successfully with selection bias, since outcomes prior to secondary school appeared to be affected by secondary school attendance (see Jones et al., 2018 for more details)[20]. We test our approach on maths scores and BMI at age 7. We would expect outcomes prior to secondary school not to be affected by attendance to a particular type of secondary school if our approach successfully isolates treatment effect from other confounders. Manning and Pischke (2006) implemented their check on age 11 outcomes, while we prefer age 7 outcomes since we use age 11 maths, reading and general ability scores to construct our ability index. Since the index constitutes the running variable for our RDD, we would expect a significant relationship with age 11 maths scores purely because of how we construct our variable. We find that age 7 maths scores are not affected by the discontinuity in grammar attendance, while age 16 scores are, a finding that holds both with and without CCT bias correction, and in the reduced form equation (see Table 3.10). This result further increases our confidence that the identification strategy is correctly isolating the effect of grammar attendance. Neither age 7 nor age 16 BMI is affected by the discontinuity, a finding that is not surprising given that this and other health outcomes were not found to be affected in our main specifications either.

A final point on robustness concerns the changing landscape of secondary schools in the 1960s and 1970s. When NCDS pupils went to secondary school, not all schools were either grammar or secondary modern. Since 1965, LEAs had been encouraged to transition to a mixed-ability system, establishing 'comprehensive' secondary schools that catered for all abilities. The Circular 10/65, promoting this move, was issued four years before NCDS cohort members took the 11-plus, and therefore there were large areas where selection was still in place. Our analysis focuses on pupils from largely selective areas: the average percentage of LEA-level comprehensive pupils for the LEAs included in our sample is 25% against 41% of the total NCDS sample, illustrating still a high degree of selectivity in our sample, in spite of the reform. As a robustness check, we also repeat our

[19] An exception is represented by the dummy for benefits recipients, which is negatively associated to grammar when excluding observations closest to the threshold, although this effect is partly offset by the positive and significant coefficients of the distance variable and its interaction with the threshold dummy

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

Table 3.10: Falsification test with outcomes prior to secondary school and pre-selected bandwidth.

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

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

main analysis including an indicator for degree of LEA selectivity, expressed by percentage of LEA-level comprehensive pupils. This is obtained from the 1971 edition of the Comprehensive School Committee Journal. Including this indicator does not affect our findings (results available on request).

3.6 Mechanism analysis

Finally, we attempt to unpack treatment effect by testing whether specific attributes of grammar schools lead to more favourable outcomes for grammar pupils. In the introduction, we mentioned three types of mechanisms, namely peers' ability and characteristics, teacher and resource quality, and curriculum. The data on school attributes is based on the school questionnaire that is part of the age 16 wave of NCDS. Peer ability is measured by looking at percentages of pupils taking either only GCEs (General Certificate of Education, equivalent to A-levels, of higher academic value) or only CSEs (Certificate of Secondary Education, of lower academic value, not requiring completion of a full standard qualification)[21]. To define peer environment, we also take an indicator for whether the school was single sex. For teaching quality, we construct a variable indicating whether teachers get any training for career guidance, and the percentage of teachers who left the school in the previous year. As a proxy of school resources, we define a binary indicator for whether the school lacks any facilities (including library, science labs, sports facilities and other). Lastly, we record weekly hours of English and Maths for each child.

The exercise is mainly exploratory in nature, since we do not provide a breakdown of the components of the full treatment effect. Rather, we provide a reduced-form evaluation of the extent to which each mechanism might explain the long-term effect of a grammar school education. We reproduce the main RDD analysis, substituting treatment in the first stage with each of the mechanisms of interest in turn. Since each characteristic is highly correlated to school type, we expect the cut-off for grammar attendance to also be predictive of such characteristics. We adopt the following two-stage empirical specification:

$$M_i = \pi_0 + \pi_1 T_i + \pi_2 D_i + \pi_3 D_i \times T_i + X_i' \kappa + \upsilon_i^M \tag{3.5}$$

$$Y_i = \mu_0 + \mu_1 \hat{M}_i + \mu_2 D_i + \mu_3 D_i \times T_i + X_i' \zeta + \epsilon_i^M, \tag{3.6}$$

In the second stage, we isolate the effect of each proposed mechanism M_i on the outcomes. Given that each M_i predicted by Equation (3.5) is highly correlated with the grammar indicator G_i, μ_1 can be read as an indirect assessment of the proportion of the effect of grammar on the outcomes that is explained by M_i, for the population of compliers.

Summary characteristics for proposed mechanisms by school type are shown in Table 3.11. In order to apply the RDD procedure as above, we dichotomise all variables before the analysis. We transform the variables for school peers' ability, teachers leaving and hours of English and Maths into binary variables taking value 1 if above the median for the relevant variable, 0 otherwise. Grammar pupils are twice as likely to attend a school with a high (above NCDS median) share of girls taking only full GCE qualifications, indicative of higher pupil ability[22]. On the other hand, secondary modern pupils are twice as likely as grammar ones to attend

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

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

Table 3.11: Descriptive statistics of mechanisms by secondary school attended.

	Grammar		**Sec. modern**			
	Mean	s.d.	Mean	s.d.	Min	Max
Peers						
Above median % girls taking GCE only	0.95	0.22	0.34	0.47	0	1
Above median% girls taking CSE only	0.31	0.46	0.76	0.43	0	1
Single sex	0.68	0.47	0.25	0.44	0	1
Teaching and resources						
Teachers get career guidance training	0.85	0.35	0.81	0.39	0	1
Above median % teachers left last year	0.47	0.50	0.52	0.50	0	1
School lacks facilities	0.49	0.50	0.58	0.49	0	1
Curriculum						
Above median hrs/week English	0.27	0.44	0.59	0.49	0	1
Above median hrs/week Maths	0.25	0.43	0.46	0.50	0	1
Observations	1160		2288			

Source: NCDS wave 3.

a school displaying a high (above NCDS median) proportion of girls taking the lower CSE qualification. A significantly higher proportion of grammar schools is single sex. Grammar schools display a slightly higher chance that teachers receive training in career guidance for their students, a lower probability of being above the median for proportion of teachers leaving the school and a lower chance that the school lacks facilities. Finally, average weekly hours of English and Maths are higher for secondary modern pupils, so they are more likely than grammar to be above the median for teaching hours. While hours are in principle a good proxy for school curriculum, these numbers may hide the real quality of the teaching and content, as well as overlooking the total actual number of hours of academic study in the two types of school. For this reason, in the following analysis, we only focus on peers and teaching and resources as possible mechanisms.

Table 3.12 shows results for the mechanism analysis, with MSE-optimal bandwidths. Given that we found a significant effect of grammar on education outcomes only, it makes sense to focus on these for the mechanism analysis[23]. Peer ability is positively linked to both educational outcomes considered. Attending a school with a proportion of GCE-takers above the median increases the probability of achieving any A-levels by 42 percentage points and a degree by 25 percentage points ($p < 0.01$ and $p < 0.05$). Being in a school with above median CSE-takers, not requiring completion of the school cycle, is negatively associated with both A-levels and degree, but the association is only significant for A-levels ($p < 0.01$). Single sex schools are also linked to better performance, increasing the chances of A-levels and a university degree by 68 and 35 percentage points

[23]Full results for all outcomes available in Appendix Table B11, and results with CCT bias correction available in Appendix Table B12

respectively ($p < 0.05$). The bias-corrected specification confirms these findings, with same coefficient sign and very similar magnitudes. Additionally, in the bias-corrected specification only, being in a school with a proportion of teachers leaving above the median significantly reduces the chances of A-levels and a degree by approximately 60 percentage points ($p < 0.1$). When bootstrapping standard errors to account for potential noise in the threshold estimation procedure, all the coefficients become insignificant at standard significance levels. However, the mechanism analysis is still informative, as it points towards an effect of some measures of peer ability and teaching resources on education outcomes, while it does not highlight any role of school facilities or specific teachers' training[24]. Unfortunately, little information was available on the quality of teaching and academic content of school curriculum, which remain important channels to consider.

3.7 Discussion

Results indicate that pupils who miss out on grammar school places within the selective system mainly miss out on higher chances to attain better educational qualifications. We find a significant effect of grammar attendance on A-levels across all specifications, while the effect on university degree becomes insignificant when we account for noise in the location of the threshold by bootstrapping standard errors. The heterogeneity analysis indicates that grammar is likely to only improve educational outcomes for pupils of high SES, or whose education parents are highly interested in. Moreover, the effect on education is larger for boys compared to girls, a finding that corroborates previous evidence (Del Bono and Clark, 2016). When isolating its effect from confounders, grammar attendance is not a significant predictor of labour market outcomes, health or risk of developing illness, compared to attending a non-selective school within the selective system. Raw differences observed in Table 3.1 are largely explained by background, and become insignificant around the threshold for grammar school entrance, where ability and other background factors are more homogeneous. As observed above, in most cases, the sign of the insignificant grammar coefficient is as expected, and Figure 3.3 suggests that significance of the grammar indicator is likely to increase as we move away from the threshold, where confounders would bias the estimate. We additionally note that OLS results in Table 3.2, for the specification adding controls, are surprisingly similar to those obtained via the RDD approach. This may suggest that controlling for observed characteristics

[24]The coefficients for these two mechanisms are larger than 1, and therefore outside the range expected. We interpret this as showing that these indicators are not picking up the effect of grammar as intended.

Table 3.12: Mechanisms: local polynomial regressions with each channel as the treatment variable and pre-selected bandwidth.

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

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

already isolates the effect of grammar to a reasonably good extent, or that, in our sample, heterogeneity in observables is relatively low.

The null result may be somewhat more surprising for wages, also given that grammar is expected to increase the probability of achieving higher educational qualifications, which should translate into better wages. A possible explanation for this is that grammar pupils scoring just above the threshold are those who are least likely to go to university, being of lower ability than the average grammar pupil. Thus, their wage outcomes may end up being as good as those for pupils ranked of highest ability among those completing secondary modern, especially given that the latter provided a more practical education that could offer a streamlined way into specific jobs (e.g. well-paid blue-collar occupations, administrative positions). An alternative explanation for the insignificant coefficient could be that the sample is too small to yield enough precision in the coefficient estimate. However, the first interpretation would also be consistent with another study using NCDS by Brunello and Rocco (2017). They find that for pupils with "lower" qualifications (approximately 11-12 years of schooling), wages are initially higher in the short-term for individuals with a vocational qualification compared to those with an academic qualification, and that this pattern is reversed by age 50. Linking this to our findings, it might be that pupils who make it into grammar and do not achieve higher qualifications experience a wage disadvantage to start with, ending up with similar wages to secondary modern pupils at age 33, as the wage profiles start to overlap. For "higher" qualifications (approximately 14-15 years of schooling), Brunello and Rocco (2017) find no significant differences in wage profiles by education type over the life-course.

Our evidence corroborates other studies that find significant and positive effects of higher quality schools on educational outcomes within a selective system, while not necessarily for labour market or health outcomes. In their study of the German context, Dustmann et al. (2017) link the lack of an effect of track assignment on education and labour market outcomes to the possibility of switching tracks in later grades. In the English system this was rarely the case, but we similarly find that grammar school attendance alone cannot explain differences in most individual outcomes, except for those directly related to educational attainment. Similar results are also found in the British context by Clark (2010) and Del Bono and Clark (2016), in the smaller areas of Yorkshire and Aberdeen respectively. Importantly, since we are estimating a local effect at the threshold, we might be missing potentially larger effects at other points of the ability distribution. For instance, grammar attendance could have a larger beneficial effect on pupils at the top of the distribution. As another example, Basu et al. (2018) find that the transition from a selective to a mixed-ability schooling system in the UK had a negative impact on smoking for individuals with lower non-cognitive ability only.

The advantage of the present paper is that we focus on the portion of the ability distribution comprising pupils who are most likely to be affected by additional grammar school places. We would expect grammar attendance to be a significant and large predictor of educational attainment for this group, since secondary modern schools at the time were geared towards vocational professions, with little or no emphasis on A-levels or higher education. This point deserves our attention. Examining the effect of grammar attendance for a cohort born in 1958 allows us to explore individuals' life trajectories over an unusually long period of time, but this also means that the school landscape has changed quite considerably since. In particular, obtaining A-levels and a university degree is now more common than it used to be, and grammar schools are not the only public institutions offering a more academic education. For instance, participation rates in higher education by age 21 are now just under 50%, compared to just above 10% in 1980 (Bolton, 2012; Mayhew et al., 2004; UK Department for Education, 2019b). Given these changes, we acknowledge that the educational advantage of grammar observed in the 1958 generation may not apply for current young generations of 11-plus takers.

The first £50 million round of the Selective Schools Expansion Fund announced in 2018 created 2,700 new grammar school places in 16 schools for the 2019/2020 academic year, and a further one has been announced in 2019 (UK Department for Education, 2019b). With 163 grammar schools present in England at the time of writing, this means that approximately 1 in every 10 grammar schools was allowed to expand. In judging the potential of this programme to reward talent rather than background, two key considerations to be made concern who can access grammar schools and what impact they can make. Previous literature has answered the first question, showing that pupils from privileged backgrounds are up to 45 percentage points more likely to attend grammar than pupils from deprived ones (Andrews et al., 2016; Burgess et al., 2018). Addressing the second issue, our findings show that past generations of grammar pupils have benefitted from grammar school attendance only in terms of a higher probability of attaining A-levels and possibly higher chances of a university degree. Moreover, we have shown that this effect was likely to apply only to pupils of high socio-economic status or whose family was highly supportive of their education. For this group, other measures of long-term human capital and health are to be linked to early background factors, such as cognitive and non-cognitive skills prior to secondary school, which remain highly predictive of the outcomes. Our findings are based on the lives of individuals who faced very different circumstances from current generations of pupils. However, this historical evidence may still be relevant, since the current expansion policy is likely to affect the intake and resources of both selective and non-selective schools in selective areas, essentially bringing back many aspects of the older system. On another note, due to the nature of the identification strategy, we did not explore the implications of the policy for pupils who are very far below the threshold, which may shed light on other important consequences of an expansion in selective school places.

Finally, we acknowledge some limitations of our study. First, sample sizes are relatively small, never above 2800 individuals, and by default even smaller within the bandwidths for LATE estimation. However, it is rare to find such detailed long-term data, and samples have been particularly small in the literature of reference in the UK (Burgess et al., 2019; Clark, 2010; Del Bono and Clark, 2016). Second, the number of outcomes surveyed may be a concern due to the issue of multiple hypothesis testing, by which the statistical probability of finding one spuriously significant coefficient when using a 5% significance level is $P = 1 - (1 - 0.05)^{10}$, approximately 40%. If we were to implement a Bonferroni correction to overcome this problem, we would have to lower the significance level to $0.005 = 0.05/n$, with $n = 10$ for the number of outcomes surveyed. We decide against this, given the small working sample, and instead opt to check the significance of grammar for educational outcomes against several robustness checks, all corroborating the main result. Third, we recognise that the missing information on actual LEA pass marks and test scores introduce imprecision in our estimates. In the absence of this information, we are still able to deliver a strong first stage in a novel way that stands up to robustness checks, thus contributing with our piece of work to expanding the empirical literature using structural breaks as discontinuities for RDD.

3.8 Conclusion

We have provided an empirical investigation of the long-term effects of grammar school attendance on human capital with a quasi-experimental methodology, exploiting a discontinuity in the probability of admission and building a novel strategy for threshold estimation from limited information. We offer a contribution to the body of research informing educational policy-makers on the effects of selective schools as means to tailor school quality to student ability. We conclude that the marginal student admitted to grammar school in the 1960s did not benefit in terms of long-term human capital accumulation, with the exception of the direct positive effect on education outcomes, which are conditional on having a favourable background. A more prominent role in explaining raw differences by type of school might be played by the child's cognitive and non-cognitive skills, and by parental support and socioeconomic background. Further research could help assess the overall impact of the grammar school system on pupils of all abilities, but we anticipate that the large role played by background characteristics will persist.

Chapter 4

Birth order effects on risky behaviours and non-cognitive skills in adolescence

Birth order effects on risky behaviours and non-cognitive skills in adolescence

Chiara Pastore*

Abstract

The adoption of risky behaviours in adolescence can affect individuals for a lifetime, and the household environment may play a large role in determining or deterring these behaviours. I investigate the effect of birth order on unhealthy and risky behaviours in adolescence, such as smoking, onset of alcohol drinking, junk food consumption, sedentary behaviour, drug use and skipping school. Birth order differences in adolescent non-cognitive skills are also examined, as a complementary explanation of why birth order effects may arise. Using a mother fixed-effect strategy to account for the endogeneity of fertility decisions and data from a panel of UK households, I find that later birth order is linked to a higher probability of engaging in early drinking, drug use and skipping school, and to lower non-cognitive skills, with some heterogeneity by sex and socio-economic status. Differences in parental investments and the influence of older siblings explain part of the observed birth order effects.

Keywords Birth order, risky behaviours, non-cognitive skills, sibling dynamics, parental investments.

JEL I12, J13, J24

*Department of Economics and Related Studies, University of York, United Kingdom.

4.1 Introduction

Are you significantly more likely to smoke, drink heavily and try drugs than your older brother or sister? Evidence shows that birth order affects individual outcomes in several domains, including education, cognitive and non-cognitive skills, earnings, adult health and delinquency (Black et al., 2005; Black et al., 2016; Black et al., 2018; Breining et al., 2017; de Haan, 2010; Hotz and Pantano, 2015; Lehmann et al., 2018; Pavan, 2016). Generally, findings go in favour of a positive link between being born earlier and higher human capital[1]. Most of this evidence analyses outcomes in either childhood or adulthood, while far less is known about birth order effects in adolescence. The physical, neurological, psychological and social changes experienced by adolescents as they transition to adulthood may result in a higher propensity to take risks, higher vulnerability to peer pressure and more rebellious attitudes. This may lead to the adoption of risky behaviours such as smoking, binge drinking and substance abuse at a young age, thus affecting an individual's long-term development, and representing a high cost for the individual as much as for society (Biglan et al., 2004; World Health Organization, 2017).

Recognising the importance of this key stage of life, this paper investigates the role of birth order in explaining differences in adolescent unhealthy and risky behaviours, looking at within-family differences to hold the background environment constant. The present study also estimates the effect of birth order on non-cognitive skills. Early non-cognitive skills are known to affect behaviour and several future outcomes, and therefore this analysis offers further insight on the role of birth order as a determinant of lifetime human capital (Kautz et al., 2014). Finally, the rest of the paper examines the role of parental investments and sibling interactions as channels to understand birth order differences in adolescent risky behaviours and non-cognitive skills.

The main difficulty in identifying the effect of birth order stems from the endogeneity of fertility decisions, since they are not independent from other mother and family characteristics. For instance, differences in family background can confound the effect of birth order, if larger families are fundamentally different from smaller ones. This seems plausible, since fertility is often related to cultural and socio-economic factors that are likely to affect a range of offspring outcomes. In order to overcome this problem, I estimate the effect of birth order with a mother fixed-effect strategy, using data on all siblings aged between 10 and 15 from Understanding Society, the UK household panel study.

[1] Throughout the paper, born earlier refers to lower birth order (e.g. first child), while born later refers to higher birth order (e.g. second, third and fourth child).

The present study contributes to both the literature on birth order effects and to that on the determinants of risk-taking in adolescence. First, it offers robust estimates of the association between birth order and six dimensions of risky adolescent behaviour, controlling for family-level unobserved confounders with the fixed-effect strategy. Second, the rich data allows measurement of rarely observed predictors of behaviours in adolescence, including individual awareness of the riskiness of unhealthy behaviours and non-cognitive skills measured via the Strengths and Difficulties Questionnaire, a high-quality validated instrument developed for accurate psychological assessment of children and adolescents. Thirdly, the household dimension of the dataset allows for a wider exploration of the channels for birth order effects. In addition to analysing the role of parental investments as in current literature, I explore how an older sibling's uptake of a risky behaviour may affect the younger sibling's decision to engage in the same behaviour herself. Lastly, I offer credible evidence on the robustness of the estimates. In addition to checking that results hold across different specifications, I test for optimal stopping behaviour by parents, to rule out that the outcomes of earlier-born children affect subsequent parental fertility decisions, which would bias the estimate of birth order effects.

I find that birth order accounts for differences in adolescent risk-taking. Being born later is associated with earlier onset of alcohol drinking and a higher probability of trying illegal drugs and skipping school. Moreover, using a psychometric instrument, I find that being born later decreases young people's non-cognitive skills. The magnitude of these associations is reasonably large when comparing it to average outcomes for first born or to gender differences. Parental investments and sibling interaction patterns are shown to account for some of the differences in risky behaviours and non-cognitive skills by birth order. Thus, the implications for public policy concern ways to equalise the costs of parental investments across all children and to incentivise positive peer effects to prevent the uptake of risky behaviours in adolescence.

The rest of the paper develops as follows. Section 4.2 illustrates previous related literature and offers a rationale for expecting birth order effects on risk-taking in adolescence. Section 4.3 presents the methodology and Section 4.4 the data used. Section 4.5 summarises the main results, followed by an assessment of heterogeneity of treatment effect. I then analyse the role of parental investments and sibling interactions as mediators for birth order effects, and implement a final robustness check of the approach. Finally, results are discussed in Section 4.6, followed by concluding remarks in Section 4.7.

4.2 Birth order effects and human capital

4.2.1 Foundations and existing evidence

Scientific interest for the role of birth order as a determinant of individual outcomes dates back to the early 1900s, with psychologist Alfred Adler's (1927) theories. In economics, the rationale for birth order effects has been formalised as a quantity-quality tradeoff in fertility decisions, first modelled by Becker (1960) and Becker and Lewis (1973). In the model, parents choose quantity of children to maximise their utility function, subject to budget and household production constraints. Since resources are limited, the relative cost of investing in children is higher for higher quantities, determining lower average parental investments for larger family sizes, and consequently lower investments for later born at any given age. Behrman and Taubman (1986) are among the first to provide an empirical application of a similar model to US data, finding that higher birth order is negatively correlated to education and earnings.

Several other empirical papers in economics have then estimated the effect of birth order on a variety of outcomes, with particular attention to the endogeneity of fertility decisions, and to isolating the effect of birth order from that of family size. Most recent studies have adopted instrumental variable (IV) or fixed-effect strategies[2]. Presence of a twin birth in the family, sex ratio in the sibship or same-sex composition of the first two births have all been used as IVs for family size, as a source of exogenous variation in the parents' decision of having a third child, thus isolating it from birth order effects (Bagger et al., 2018; Black et al., 2005; Brinch et al., 2017; Conley and Glauber, 2006; de Haan, 2010; Grinberg, 2015; Kumar, 2016). The main drawback of this approach is that it looks at a local average treatment effect for compliers, who are families of a specific type (e.g. families with twins or three children), so the generalisability of its findings is limited. The fixed-effect specification on the other hand, accounts for family size by netting out its effect together with all other characteristics that are fixed between siblings. Studies using mother fixed effects to estimate birth order effects often exploit large Scandinavian administrative data or US survey data that collect information on all siblings in a family. Most of these studies focus on education outcomes, although recently there has been a growing interest for birth order effects on skill production, infant and adult health and delinquency (Barclay and Kolk, 2017; Barclay and Myrskylä, 2014; Björkegren and Svaleryd, 2017; Black et al., 2016; Black et al., 2018; Breining et al., 2017; Brenøe and Molitor, 2018; Lehmann et al., 2018; Pavan, 2016). Generally, studies find a negative and significant

[2]See Appendix Table C12 for a summary of relevant papers.

relationship between higher birth order and most outcomes[3].

The following are the four most relevant studies for the present paper, although only the first three use mother fixed effects. Black et al. (2016) analyse birth order effects on adult health behaviours and health, with administrative data from Norway. They find that late born display higher smoking rates and lower self-assessed physical and mental health, although they score better than first born on body mass index and other biomarkers. Black et al. (2018) find that birth order negatively affects adult non-cognitive skills in a sample of Swedish men examined as part of the military enlistment. Later-born men are also less likely to be in jobs that require leadership and they score lower in all Big 5 personality traits. Both cited studies find lower parental investments for higher birth order children, including behaviours such as smoking during pregnancy, breastfeeding and help with homework. A similar conclusion is reached by Lehmann et al. (2018), who use a cohort of children from the US (NLSY-C) to assess birth order effects on children cognitive and non-cognitive skills, as well as educational outcomes in adulthood. They find that lower parental investments, measured as antenatal care, behaviour in pregnancy, breastfeeding and cognitive stimulation, can account for the negative effect of higher birth order on cognitive skills in childhood, but find no effect on average childhood non-cognitive outcomes. Finally, Argys et al. (2006) are the only other economic study looking at adolescent risky behaviours. Using NLSY79, a longitudinal study of adolescents from the US, they find a positive association between higher birth order and smoking, use of marijuana, alcohol consumption and earlier initiation of sexual behaviour in adolescence. They estimate OLS regressions with birth order expressed by a dummy variable for having any older siblings. In the present paper, I use a fixed-effect strategy, netting out the effect of family size and other family-specific factors that are constant across groups of siblings. The strategy thus allows for a more robust estimation of the association between birth order, risky behaviours and non-cognitive skills for adolescents, a population that has not been examined before in studies using fixed-effect strategies. Institutionally, the ages considered correspond to the period going from the end of primary school to the end of compulsory schooling, covering a key stage of individual development.

[3]This relationship is inverted in developing countries, as demonstrated by studies on Nigeria, Brazil, Ecuador and India. An explanation offered for the positive effect of higher birth order on education, child labour and other outcomes in these contexts is that lower-order children are more likely to be born when parents have fewer resources, sometimes in their teens and unable to provide for them appropriately (de Haan, 2010; Emerson and Souza, 2008; Kumar, 2016; Tenikué and Tequame, 2017).

4.2.2 Rationale for birth order effects on risky behaviours and non-cognitive skills

When thinking about risky behaviours and non-cognitive skills, it is not clear, a priori, whether birth order effects should be positive or negative. I focus on two main mechanisms: parental behaviours and sibling interactions.

Parents may spend more time with earlier-born children because their job responsibilities early on in their career allow it or because the utility from first-time parenthood is higher, and thus teach them healthier behaviours. Earlier-born children generally have parents who are younger at birth, and therefore may have more energies to invest in their children. However, later born have more experienced parents, who may be more efficient in transferring important messages about health behaviours. Similar considerations also hold for non-cognitive skill production, of which parental investments are a large component (Cunha et al., 2010; Moroni et al., 2019). In psychology, Blake's (1981) resource dilution hypothesis expresses the idea that parental resources for each additional child will decrease, as they are shared among a larger number of children. Price (2008) finds evidence of resource dilution by showing that, at the same age, first born in the US receive more parental quality time than second born. Further, Hotz and Pantano (2015) show that parents are stricter with first-born than later-born children, while Pavan (2016) finds that parental behaviours account for between 20% and 45% of the cognitive skill gap by birth order. Another point concerns biological differences by birth order. Studies in the medical literature have shown that the placenta is more conductive of nutrients after the first pregnancy (Juntunen et al., 1997; Khong et al., 2003; Wilcox et al., 1996 cited in Black et al., 2018), and there are studies in economics showing that first born have worse health at birth than their younger siblings (Björkegren and Svaleryd, 2017; Brenøe and Molitor, 2018). While such biological differences should not in principle affect risk-taking directly, if first born have worse health in infancy, parents may be particularly protective of them, and invest more in transferring healthy habits to them than to later born.

Turning to sibling interactions, if earlier born take care of their younger siblings, they may develop a higher sense of responsibility and be less likely to engage in risk-taking behaviours. In terms of non-cognitive skills, first born may become more conscientious and responsive to parental expectations, while later born may be extraverted and competitive, growing up trying to get parental attention. This idea is consistent with Sulloway (1996), who proposes a 'family dynamic model' in the psychology literature, where siblings compete for parental attention, and shape their personalities accordingly, often differentiating themselves for this purpose. In terms of behavioural outcomes, later born may learn by imitation, by observing and reproducing their older siblings' behaviours. This mechanism may explain an earlier uptake of a risky behaviour by later born children, which could be more damaging to their health if they start at a younger age. In line with this, Harakeh et al. (2007) show that friends and older siblings influence younger siblings' smoking behaviour in adolescence. Secondly, older siblings may affect younger ones by increasing the information available to them about the riskiness of certain behaviours, although the direction of this effect is also unknown. Thanks to the household-level information, in this paper I can test for birth order effects on this dimension, by looking at whether older siblings' behaviour affects younger children's' behaviour and beliefs about risky behaviours.

4.3 Empirical methods

The set of risky behaviours and non-cognitive skills Y are assumed to be the product of a function $Y(B, \mathbf{F}, \mathbf{X})$, where B denotes birth order, $\mathbf{F}$ is a vector of family characteristics that are constant across siblings, including family size, and $\mathbf{X}$ a vector of individual-specific characteristics. The simplest method to look at the association between birth order and the outcomes consists of running ordinary least squares (OLS) regressions. In a first specification, birth order is expressed as a categorical variable B_{ij} for individual i and mother j, taking values from 1 to 4 for each corresponding birth order. Thus, keeping $B_{ij} = 1$ as base category, I estimate:

$$Y_{ij} = \beta_1 + \sum_{k=2}^{4} \beta_k 1(B_{ij} = k) + \mathbf{F}'_{ij}\gamma + \mathbf{X}'_{ij}\delta + \epsilon_{ij}. \tag{4.1}$$

In Equation (4.1), β_1 is a constant and ϵ_{ij} an idiosyncratic error term. The choice of included covariates is a parsimonious set of variables that reflect family and individual background. Family characteristics $\mathbf{F}_{ij}$ include dummies for sibship size, mother's highest qualification, family gross monthly income, an indicator for lone-parent families and dummies for region of residence. The individual characteristics vector $\mathbf{X}_{ij}$ includes sex, an indicator for nonwhite ethnicity, year of birth (to control for cohort effect), dummies for age at which response was recorded (to control for variations in behaviour due to age), and mother's age when individual was born[4]. Each β_k estimated by OLS yields the association between being kth-born and the outcome, compared to being a first-born child. However, the estimated β_k is likely to be biased, due to the possible presence of unobservable traits correlated with birth order, causing $cov(B_{ij}, \epsilon_{ij}) \neq 0$. In my

[4]Several studies include mother's year of birth, but here it is excluded to avoid perfect multicollinearity, since Mother's YOB = Child's YOB - Mother's age at birth.

preferred specification, I include mother fixed effects ϕ_j to deal with this bias.

$$Y_{ij} = \beta_1' + \sum_{k=2}^{4} \beta_k^{FE} 1(B_{ij} = k) + \phi_j + \mathbf{X}_{ij}' \delta^{FE} + \epsilon_{ij}'. \tag{4.2}$$

The fixed effect (or within) estimator nets out the family-specific observable and unobservable characteristics by first demeaning the data, using mean values of all variables calculated for individuals from the same mother, and then estimating the resulting equation by pooled OLS (Woolridge, 2013):

$$\ddot{Y}_{ij} = \sum_{k=2}^{4} \beta_k^{FE} \ddot{B}_{ij} + \ddot{\mathbf{X}}_{ij}' \delta^{FE} + \ddot{\epsilon}_{ij}, \tag{4.3}$$

where the demeaned variables are such that $\ddot{Y}_{ij} = Y_{ij} - \bar{Y}_j$, and $\bar{Y}_j$ is the average Y for all individuals with mother j. When estimating Equation (4.3), all mother-specific characteristics ϕ_j holding constant across siblings are removed[5]. Under the assumption that the mix of genetic and other unobservable characteristics is randomly assigned among siblings at birth, any residual systematic difference observed by birth order reflects in principle a pure effect of birth order, expressed by β_k^{FE}. An important exception to this statement is represented by unobserved circumstances that may systematically affect later born differently from earlier born. To give a concrete example, take the hypothetical case that parental divorce always affects later born more severely, since on average they are younger when they experience it. Then, the β_k^{FE} estimate would be confounded by the effect of experiencing parents' divorce at an earlier age[6]. For this reason, I refrain from claiming causality of β_k^{FE} outright, and instead argue for a robust correlation, showing it holds across a variety of specifications and checks.

I further explore other aspects of the effect of birth order within the family, by varying the definition of the treatment variable. The first alternative specification uses only one indicator for having any older siblings instead of the birth order dummies. The second one distinguishes between having older siblings of a different sex, and having older siblings of the same sex, compared to being first born, in order to look at whether effects of birth order differ by sibship sex com-

[5]This approach is equivalent to fixed effect estimation on panel data, with mothers as the N dimension, and the number of children for each mother as the T. The resulting structure resembles an unbalanced panel. For an unbiased FE estimator in the presence of an unbalanced panel, attrition should be random, as detailed in Wooldridge (2010). In this scenario, the different number of observations T for each N is not due to missing data, but to different numbers of children, which are likely correlated to other mother characteristics. A way to get around this problem is to estimate the models separately by family size, so that all analysis is implemented on a balanced panel. Results are qualitatively the same with this approach. See Appendix Section C.3 for further discussion.

[6]Results are qualitatively the same when excluding non-cohabiting parents from the sample.

position. More alternative definitions of birth order and respective results can be found in Appendix Section C.4.

4.4 Data

The data are from the first eight waves of Understanding Society, the UK longitudinal household panel study started in 2009 with approximately 40,000 households and repeated annually since, with some attrition in the process (Lynn and Borkowska, 2018)[7]. The household feature of the study permits observation of all individuals residing in the same household at a given point in time, the main focus being children in the household. All outcomes are taken from questions in the Young Person element of the survey, administered to all individuals in the household aged between 10 and 15[8]. The total sample of included individuals amounts to N=$10,132$, although samples included for each outcome vary, due to missing data items.

4.4.1 Birth order and sibship

The analysis focuses on families where I observe at least two siblings born from the same mother, which is central for the fixed effects strategy. I thus exclude 1,016 households with one child, while households with twins are excluded to avoid confusion over birth order classification, leaving out a further 327 individuals. I also exclude adopted and foster children and limit the sample to families of at most four children, to ensure similarity and large enough sample sizes. The birth order variable is constructed from information on all natural children had by a mother and their year of birth[9]. Of the total 10,132, 40.9% are first born, 38.8% are second born, 15.8% are third born and 4.5% are fourth born. A variable for sibship size is similarly constructed, including all siblings from the same mother, including those not living in the household at the time of the survey.

Table 4.1 shows descriptive statistics by birth order. The proportion of girls is constant across birth order, and average birth year is between 1996 and 1997. First and second born present similar percentages of non-white children, roughly

[7]Attrition is only a problem for the estimates if the probability of observing younger children in the survey is correlated with other household or child characteristics. For example, if poorer families are more likely to drop out, I may not observe their later born, and therefore results would only hold for the sample of wealthier families. See Appendix Section C.3 for more details on 'unbalanced' households.

[8]More information on the survey design and fieldwork can be found at https://www.understandingsociety.ac.uk/documentation/mainstage/technical-reports.

[9]Even if not all siblings from the same mother are subsequently observed in the survey, an individual can be included in the sample as long as at least another sibling sharing the same mother completes the individual questionnaire.

Table 4.1: Descriptive statistics of background characteristics by birth order for individuals aged 10-15.

	1st born	2nd born	3rd born	4th born
Female	0.493	0.497	0.495	0.457
	(0.500)	(0.500)	(0.500)	(0.499)
Birth year	1996.6	1996.9	1996.9	1996.7
	(4.859)	(4.804)	(4.851)	(4.702)
Not white	0.258	0.235	0.276	0.313
	(0.438)	(0.424)	(0.447)	(0.464)
Sibship size	2.507	2.554	3.253	3.911
	0.681	0.697	0.502	0.299
Years from closest older sibling	.	3.647	4.289	4.098
	(.)	(2.729)	(3.378)	(3.129)
Years from eldest sibling	.	3.647	7.418	10.27
	(.)	(2.729)	(4.116)	(4.243)
Mother's age at birth	26.16	29.39	31.49	33.02
	(5.140)	(5.043)	(4.996)	(4.756)
Mother's birth year	1970.4	1967.5	1965.5	1963.7
	(6.637)	(6.381)	(6.425)	(6.265)
Mother's education				
No qualification	0.091	0.101	0.180	0.243
Lower qualification	0.074	0.092	0.119	0.127
GCSE	0.266	0.264	0.247	0.237
A-level	0.188	0.177	0.145	0.154
Non-degree higher qualification	0.154	0.161	0.144	0.121
Degree	0.226	0.206	0.164	0.119
Gross HH monthly income	3718.9	3894.7	3797.1	3699.5
	(2512.8)	(2656.4)	(2773.7)	(2651.4)
Lone parent HH	0.207	0.197	0.202	0.210
	(0.405)	(0.398)	(0.402)	(0.408)
Observations	4149	3927	1603	453

Mean values; standard deviation in parentheses

around a quarter, while this proportion rises to 30% for fourth born. Average sibship size and distance in years from the next older as well as the eldest sibling increase gradually for later-born children, and so does average mother's age at birth[10].

4.4.2 Household background

Household characteristics to be included as regressors are mother's age at birth, as well as her highest educational qualification recorded in the first survey response. Family's gross monthly income is based on household income in the previous month. Approximately 20% of households are lone parent households, defined

[10]Median distance between siblings in years is somewhat lower. The median from the closest older sibling is 3 years for all categories, while median values for years from eldest sibling are 3, 6, and 9 for 2nd, 3rd and 4th born respectively.

as households where there is only one parent living without a partner with their dependent children[11]. Region of residence is also included as a regressor, and it is one of 12 government office regions. Fewer than 2% households in the sample move region over the survey period, and therefore this is taken as a time-constant characteristic. Table 4.1 shows that mothers of later-born children are less educated on average. Average gross household monthly income, on the other hand, is highest for second-born children, and it declines later for third and fourth born. This may reflect that parents earn more as they age, but this may be reversed in larger families if parents are less educated on average or if they decide to work fewer hours to look after their children. As expected, differences in background characteristics by birth order suggest that selection bias due to observables (and potentially unobservables) is likely to be a problem in this setting.

4.4.3 Parental investments

Understanding Society contains information on several measures of parental behaviour, useful to explore the sources of birth order effects. The survey provides both information on whether the mother ever breastfed the baby, and for how long, recalled by the mother in the first wave for each child. Based on this information, I construct a variable counting how many months the child was breastfed for. Additional information on parental investment was retrieved from the child's response to the Young Person questionnaire. This includes whether the child feels their parents are interested in how they do at school; whether their parents help them with their homework; whether the child always feels supported by their family; and whether they turn to their parents first when upset[12]. Parental investments are shown to vary by birth order in Table 4.2. First-born children are most likely to have been breastfed at all and for longer, and second-born least likely. It should be noted that larger families may have different practices and this may be why the averages are higher for third- and fourth-born children. Parental interest in school as well as the probabilities of parents helping with homework, feeling supported by one's family and talking to one's parents when upset are highest for first born and decline with birth order.

[11]Results are the same if the analysis is conducted excluding lone parent households.

[12]These questions are also asked directly to parents about all their children, although the questions do not differentiate between each child. This means that the adult survey cannot capture potential variation in parental investment across different children.

Table 4.2: Descriptive statistics of parental investments by birth order.

	1st born	2nd born	3rd born	4th born
Was breastfed	0.664	0.613	0.649	0.630
	(0.473)	(0.488)	(0.478)	(0.486)
Months was breastfed for	4.669	4.162	3.912	5.001
	(6.653)	(6.334)	(5.926)	(6.468)
Parents interested in school	0.839	0.808	0.807	0.837
	(0.368)	(0.394)	(0.395)	(0.370)
Parents help with homework	0.867	0.807	0.761	0.661
	(0.340)	(0.395)	(0.427)	(0.475)
Always feels supported by family	0.800	0.778	0.766	0.748
	(0.400)	(0.415)	(0.424)	(0.435)
Talks to parents when upset	0.827	0.714	0.647	0.585
	(0.379)	(0.452)	(0.478)	(0.494)
Observations	4149	3927	1603	453

Mean values; standard deviation in parentheses

4.4.4 Outcomes

Outcomes of interest are risky behaviours and non-cognitive skills. Responses are recorded at the most recent available wave for each individual, and age at response is also recorded, to control for differences in outcomes driven by age.

Behaviours

I select the following indicators of unhealthy risky behaviours: whether the individual ever smokes cigarettes at all, age of first alcoholic drink, number of days the child has junk food weekly ("eat crisps or sweets or have fizzy drinks such as Coke or lemonade"), and times a week the individual "plays sports, does aerobics or does some other keep fit activity". This variable is recoded so it is increasing in sedentary behaviour. I also include binary indicators for having tried any illegal drugs and having played truant at (i.e. skipped) school in the last 12 months.

Table 4.3 shows that behaviours vary by birth order category. On average, second-, third- and fourth-born children are more likely to smoke, and to start drinking earlier (about 0.2 of a year, or almost 2 and a half months), compared to first born. The diet variable is coded as a dummy variable, taking value 1 if the child says they have crisps, sweets or fizzy drinks almost everyday or more, and 0 otherwise. Later-born children display a higher probability of having unhealthy diets compared to first born. Exercise patterns do not show clear differences between first and second born, while third and fourth born do less weekly exercise on average. Children are more likely to have tried illegal drugs and to have played truant at school in the last 12 months if they are not first born.

Table 4.3: Descriptive statistics of outcomes by birth order for individuals aged 10-15.

	1st born	2nd born	3rd born	4th born
Behaviours				
Smokes	0.0195	0.0244	0.0268	0.0221
	(0.138)	(0.154)	(0.162)	(0.147)
Age started drinking (years)	14.56	14.36	14.31	14.40
	(3.444)	(3.475)	(3.461)	(3.377)
Regularly eats junk food	0.411	0.444	0.458	0.455
	(0.492)	(0.497)	(0.498)	(0.499)
Exercise (times/week)				
Everyday	0.1809	0.1679	0.1609	0.1763
4 or 5	0.1657	0.1797	0.1468	0.1492
3 or 4	0.2893	0.3011	0.3067	0.2949
1 or 2	0.2579	0.2535	0.2625	0.2847
Less than 1	0.0637	0.0586	0.0753	0.0475
Never	0.0425	0.0392	0.048	0.0475
Ever tried illegal drugs	0.0296	0.0458	0.0463	0.0398
	(0.170)	(0.209)	(0.210)	(0.196)
Play truant	0.0975	0.112	0.127	0.0918
	(0.297)	(0.316)	(0.333)	(0.289)
Non-cognitive skills				
SDQ prosocial score (0-10)	7.717	7.588	7.553	7.621
	(1.824)	(1.875)	(1.891)	(1.792)
SDQ internalising score (0-20)	4.577	4.574	4.685	4.554
	(3.227)	(3.232)	(3.279)	(3.173)
SDQ externalising score (0-20)	5.870	6.168	6.274	6.086
	(3.551)	(3.569)	(3.678)	(3.405)
Observations	4149	3927	1603	453

Mean values; standard deviation in parentheses

Non-cognitive skills

Non-cognitive skills are assessed via the Strengths and Difficulties Questionnaire (SDQ) in the Young Person survey. Developed by psychiatrist Robert Goodman, the self-reported version of the questionnaire is appropriate for behavioural screening of children and adolescents (Goodman et al., 1998). The questionnaire includes 25 attributes grouped under five different categories (see Appendix Section C.2). For each attribute, children could pick "Not true", "Somewhat true" or "Certainly true" (scoring 0, 1 or 2), depending on how closely they felt the attribute applied to them. In populations at low risk of severe behavioural disorders, a three-part classification of the SDQ score is recommended (Goodman et al., 2010). The first category, prosocial behaviour, constitutes an outcome on its own, ranging from 0 to 10, and it is decreasing in behavioural problems and increasing in non-cognitive skills. Questions falling under the emotional symptoms and peer problem subcategories are incorporated into an internalising score, expressing internal problems and negative feelings such as anxiety and depression. Finally, the scores for conduct problems and hyperactivity/inattention are combined into an externalising score, expressing instead external shows of hostility, antisocial behaviour and aggression (American Psychological Association, 2018). Both internalising and externalising scores range from 0 to 20, and they are increasing in behavioural problems and decreasing in non-cognitive skills[13]. As shown in Table 4.3, average prosocial score decreases with higher birth order, indicating lower non-cognitive skills, although it increases slightly for fourth born. Internalising and externalising scores increase gradually for higher birth order, similarly indicating lower non-cognitive skills for later birth order, with the only exception of fourth born.

4.5 Results

4.5.1 Effect of birth order

Table 4.4 shows the effect of birth order for all behavioural outcomes considered. For each outcome, the first column displays OLS regression results, controlling for all covariates, while the second column includes mother fixed effects. Standard errors are clustered at the mother's level, while age and birth year are included as dummies to flexibly allow for nonlinearities in the relationship with the outcomes. Regressions for binary outcomes (smokes, regular junk food consumption, tried drugs, play truant) are estimated via linear probability models, to avoid the bias of maximum likelihood estimators based on probit models in the presence of fixed effects (Greene, 2004)[14]. All behavioural outcomes display a significant association with birth order when estimated via OLS regressions in panel A, the base category being first-born children. Later born are more likely to smoke and start drinking earlier, to consume more junk food, to have tried illegal drugs and played truant in the last 12 months. Mostly, the magnitude of the coefficient increases as birth order increases, as well as the associated standard error.

The OLS associations are confirmed by the fixed-effect specification, with the exception of smoking, junk food and sedentary behaviours. The fixed-effect coefficients are larger, but so are the standard errors. Compared to first born, age at first drink decreases by between 20% and 80% of a year for later born, corresponding roughly to 2.5-9.5 months earlier, with coefficients increasing in magnitude

[13]Using the alternative SDQ classification into prosocial and difficulty scores yields similar results across all the analysis. The difficulty score is given by the sum of internalising and externalising scores.

[14]More recent developments in the literature are using bias correction methods to deal with this problem in large samples (Cruz-Gonzalez et al., 2017; Fernández-Val and Weidner, 2016). In the present context, the parsimonious and still informative approach of a linear probability model is preferred, although using probit and ordered probit models gives very similar results.

Table 4.4: Birth order and risky behaviours

	Smokes		**Age started drinking**		**Junk food**		**Tried drugs**		**Play truant**		**Sedentary behaviour**	
Panel A												
Second child	0.0187*	0.0134	-0.1109**	-0.2638***	0.0606***	0.0246	0.0177**	0.0322**	0.0216*	0.0441**	-0.0201	-0.1632*
	(0.0077)	(0.0186)	(0.0406)	(0.0741)	(0.0138)	(0.0265)	(0.0060)	(0.0124)	(0.0084)	(0.0164)	(0.0348)	(0.0665)
Third child	0.0311*	0.0037	-0.2067**	-0.5003***	0.0986***	0.0328	0.0216*	0.0521*	0.0222	0.0634*	+0.0660	-0.2131
	(0.0124)	(0.0337)	(0.0668)	(0.1437)	(0.0221)	(0.0502)	(0.0092)	(0.0213)	(0.0136)	(0.0297)	(0.0574)	(0.1339)
Fourth child	0.0325	-0.0171	-0.3663***	-0.8760***	0.1092**	0.0298	0.0176	0.0678	-0.0035	0.0847	-0.0593	-0.2671
	(0.0215)	(0.0586)	(0.1054)	(0.2248)	(0.0380)	(0.0772)	(0.0169)	(0.0381)	(0.0211)	(0.0455)	(0.0932)	(0.2065)
Female	0.0015	0.0054	0.0149	0.0501	-0.0314*	-0.0409*	-0.0130*	-0.0188*	-0.0066	-0.0065	0.4922***	0.4567***
	(0.0071)	(0.0121)	(0.0363)	(0.0598)	(0.0126)	(0.0190)	(0.0054)	(0.0083)	(0.0074)	(0.0115)	(0.0320)	(0.0510)
Panel B												
Any older siblings	0.0206**	0.0198	-0.1281**	-0.1402*	0.0668***	0.0222	0.0183**	0.0234*	0.0214**	0.0350*	-0.0071	-0.1398*
	(0.0075)	(0.0149)	(0.0408)	(0.0653)	(0.0136)	(0.0228)	(0.0058)	(0.0106)	(0.0082)	(0.0144)	(0.0344)	(0.0560)
Female	0.0013	0.0056	0.0168	0.0573	-0.0316*	-0.0409*	-0.0130*	-0.0189*	-0.0065	-0.0069	0.4925***	0.4577***
	(0.0071)	(0.0120)	(0.0362)	(0.0596)	(0.0126)	(0.0190)	(0.0054)	(0.0083)	(0.0074)	(0.0115)	(0.0320)	(0.0510)
Panel C												
Older sib, diff sex	0.0218*	0.0250	-0.1068*	-0.1151	0.0714***	0.0279	0.0194**	0.0268*	0.0206*	0.0388*	-0.0602	-0.1798**
	(0.0091)	(0.0166)	(0.0468)	(0.0712)	(0.0161)	(0.0244)	(0.0072)	(0.0119)	(0.0097)	(0.0153)	(0.0406)	(0.0624)
Older sib, same sex	0.0196*	0.0115	-0.1474**	-0.1795*	0.0627***	0.0140	0.0173**	0.0185	0.0221*	0.0293+	0.0390	-0.0831
	(0.0087)	(0.0159)	(0.0466)	(0.0752)	(0.0157)	(0.0265)	(0.0066)	(0.0115)	(0.0094)	(0.0167)	(0.0392)	(0.0629)
Female	0.0013	0.0054	0.0163	0.0573	-0.0317*	-0.0410*	-0.0130*	-0.0188*	-0.0065	-0.0068	0.4937***	0.4574***
	(0.0071)	(0.0121)	(0.0362)	(0.0595)	(0.0126)	(0.0190)	(0.0054)	(0.0083)	(0.0074)	(0.0115)	(0.0320)	(0.0510)
Fixed effects	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4076	4076	6120	6120	6201	6201	5185	5185	6927	6927	6593	6593

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

Table 4.5: Birth order and non-cognitive skills

	SDQ Prosocial		SDQ Interalising		SDQ Externalising	
Panel A						
Second child	-0.1396**	-0.1272	0.0251	0.1423	0.5004***	0.7679***
	(0.0512)	(0.1022)	(0.0895)	(0.1775)	(0.0993)	(0.1913)
Third child	-0.0291	-0.0003	0.0480	0.4758	0.5815***	1.1530**
	(0.0828)	(0.1979)	(0.1446)	(0.3505)	(0.1642)	(0.3626)
Fourth child	0.2501+	-0.0170	-0.0795	0.3509	0.5049+	1.6557**
	(0.1336)	(0.3123)	(0.2403)	(0.5560)	(0.2612)	(0.5630)
Female	0.8476***	0.9481***	1.0813***	1.0550***	-0.6990***	-1.1085***
	(0.0459)	(0.0740)	(0.0805)	(0.1222)	(0.0893)	(0.1380)
Panel B						
Any older siblings	-0.1188*	-0.1676+	0.0274	0.0476	0.5127***	0.5678***
	(0.0502)	(0.0861)	(0.0882)	(0.1461)	(0.0981)	(0.1656)
Female	0.8455***	0.9471***	1.0820***	1.0537***	-0.6988***	-1.1189***
	(0.0459)	(0.0739)	(0.0804)	(0.1221)	(0.0893)	(0.1383)
Panel C						
Older sib, diff sex	-0.1393*	-0.1905*	0.0535	0.0808	0.5767***	0.6332***
	(0.0593)	(0.0939)	(0.1028)	(0.1559)	(0.1140)	(0.1767)
Older sib, same sex	-0.1007+	-0.1341	0.0045	-0.0011	0.4564***	0.4727*
	(0.0580)	(0.0991)	(0.1024)	(0.1722)	(0.1143)	(0.1953)
Female	0.8460***	0.9473***	1.0814***	1.0533***	-0.7002***	-1.1198***
	(0.0459)	(0.0739)	(0.0805)	(0.1221)	(0.0893)	(0.1383)
Fixed effects	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6219	6219	6212	6212	6214	6214

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

with later birth order. Being born later also increases the probability of having tried any illegal drugs, by between 3 and 6 percentage points, relative to the 3% mean for first born, and the probability of having played truant at school by between 4 and 8 percentage points, relative to a first born average of 9.8%. The fixed-effect results also indicate that second-born children engage in more weekly exercise than first born, but this relationship is not confirmed with later birth orders. Given that the association of birth order with smoking and junk food consumption disappears after introducing fixed effects, other family characteristics may account for most of the observed correlation between being born later and engaging in smoking and unhealthy diets. These could be observed, such as family size or socio-economic status, but also family-specific unobservables.

Table 4.4 also displays results with two alternative treatments in panels B and C. One is a binary indicator for having any older siblings; the other explores sex composition, with indicators reporting whether the individual has any older siblings of a different or same sex to their own, the base category still being first born. Estimates mostly confirm results from the main specification, uncovering some further information on birth order effects. Coefficients are sizeable and significant for drinking age, drugs, playing truant and sedentary behaviour. However, while for drinking age the effect is driven by children whose older siblings are of the same sex, for the other three behaviours the effect is stronger for children whose older siblings are of the opposite sex.

Table 4.5 shows the effects of birth order on the three non-cognitive skills measures in adolescence. In panel A, the OLS specification shows a significant association between being born second and lower prosocial score (i.e. lower non-cognitive skills), but this is not significant in the fixed-effect specification. Further, no significant effect of birth order is observed on internalising score. On the other hand, compared to first born, being born second, third or fourth significantly increases the SDQ externalising score, decreasing in non-cognitive skills. Again, coefficients are mostly larger in the fixed-effect specification, which may indicate that family-constant characteristics bias the effect of birth order on the externalising score towards zero, but standard errors are also larger. The effect ranges between an increase by 0.8 points (for second born) up to more than 1.6 points, relative to a mean of 5.9 for first born. Table 4.5 also displays results with the alternative definitions of birth order, in panels B and C. The effect of being born later on prosocial score is negative and now statistically significant, even when including fixed effects. The coefficient has a magnitude of roughly 0.2 points and the association is driven by children whose older siblings are of the opposite sex. Again, there is no significant effect on internalising score, while externalising score is significantly and positively affected by birth order[15].

An indicator for female is displayed in all regressions for comparison purposes. Junk food consumption, exercise levels and the probability of trying illegal drugs are significantly lower for girls than boys. For instance, the association between birth order and trying drugs in adolescence is twice as large as that of sex and trying drugs. Girls also display higher prosocial and internalising score than boys, but lower externalising score. For externalising score, being a girl offsets the negative effect of being born later on non-cognitive skills (the magnitude of the association is twice as large).

Finally, to explore the components of family fixed effects, I regress the predicted fixed-effect term from Equation (4.2) on all included family characteristics, for each of the nine outcomes. Results, available in Tables C1 and C2 in the Appendix, show that some observed characteristics are important, although for most outcomes the constant term is highly significant, meaning that a large portion of

[15]When estimating the same models separately by family size, sample size shrinks considerably and estimates are less precise, but the main results hold and coefficients maintain significance at the conventional levels.

characteristics accounting for fixed effects is unobserved.

4.5.2 Heterogeneity analysis

To relax the assumption of a constant average effect of birth order for all individuals, I conduct a heterogeneity analysis, focusing on a set of background characteristics expected to shift treatment effect. The four different sources of heterogeneity evaluated are sex and ethnicity of the child, mother's education and family income. This analysis is particularly useful if birth order effects act in opposite directions for different subgroups of the population. I thus allow for different production functions by repeating estimation separately for each pair of subgroups, for all outcomes. Subgroup pairs are girls versus boys, white versus non-white, those having a mother with a degree versus without a degree, those whose family income is above the median versus below median.

Results are displayed in Figures 4.1-4.4. When grouping individuals and estimating regressions separately by sex, it is evident that the effect of birth order is driven by boys. This is especially the case for smoking, drug use and playing truant. Moreover, the null average effect of birth order on prosocial and internalising score observed in Table 4.5 might be hiding different dynamics by sex. Boys display a negative effect of birth order on prosocial score and a positive effect on internalising score. For age at first drink and externalising behaviour, the effects overlap for the two groups. Fewer differences are found when dividing the sample by mother's education, family income and ethnicity. The effect of birth order on non-cognitive skills is slightly more pronounced and negative for children whose mother has a degree, compared to those without. A speculation is that this may be due to mothers with degrees having steeper increases in time demanded by their jobs as they progress in their career, proportionally reducing time with later children more than for mothers without a degree, who may experience fewer changes in their time commitments. Compatible with this explanation is that when grouping individuals by income, the positive effect of birth order on externalising scores is exacerbated for families whose income is above the median. Finally, birth order effects for the prosocial scores go in opposite directions for white compared to non-white children, although 95% confidence intervals overlap for the two groups for all birth order coefficients[16].

[16] In an alternative specification, I allow for an interaction between each characteristic expressed as a dummy and the birth order indicator (equal to one for female, non-white, the mother having a university degree and family income above median respectively). For the interested reader, results are shown in Appendix Tables C3-C10.

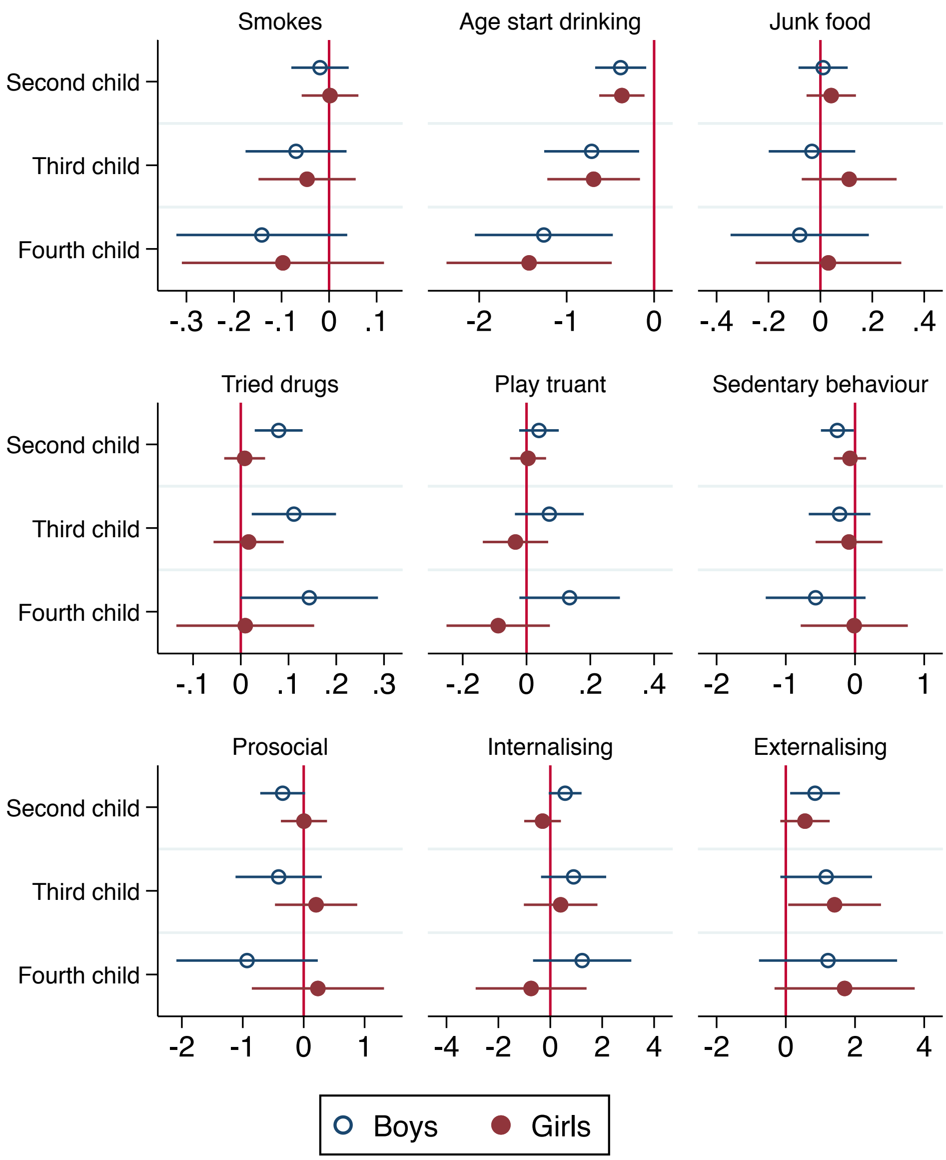


Figure 4.1: Heterogeneity of birth order effects by sex. Markers represent point-estimates and lines illustrate respective 95% confidence intervals.

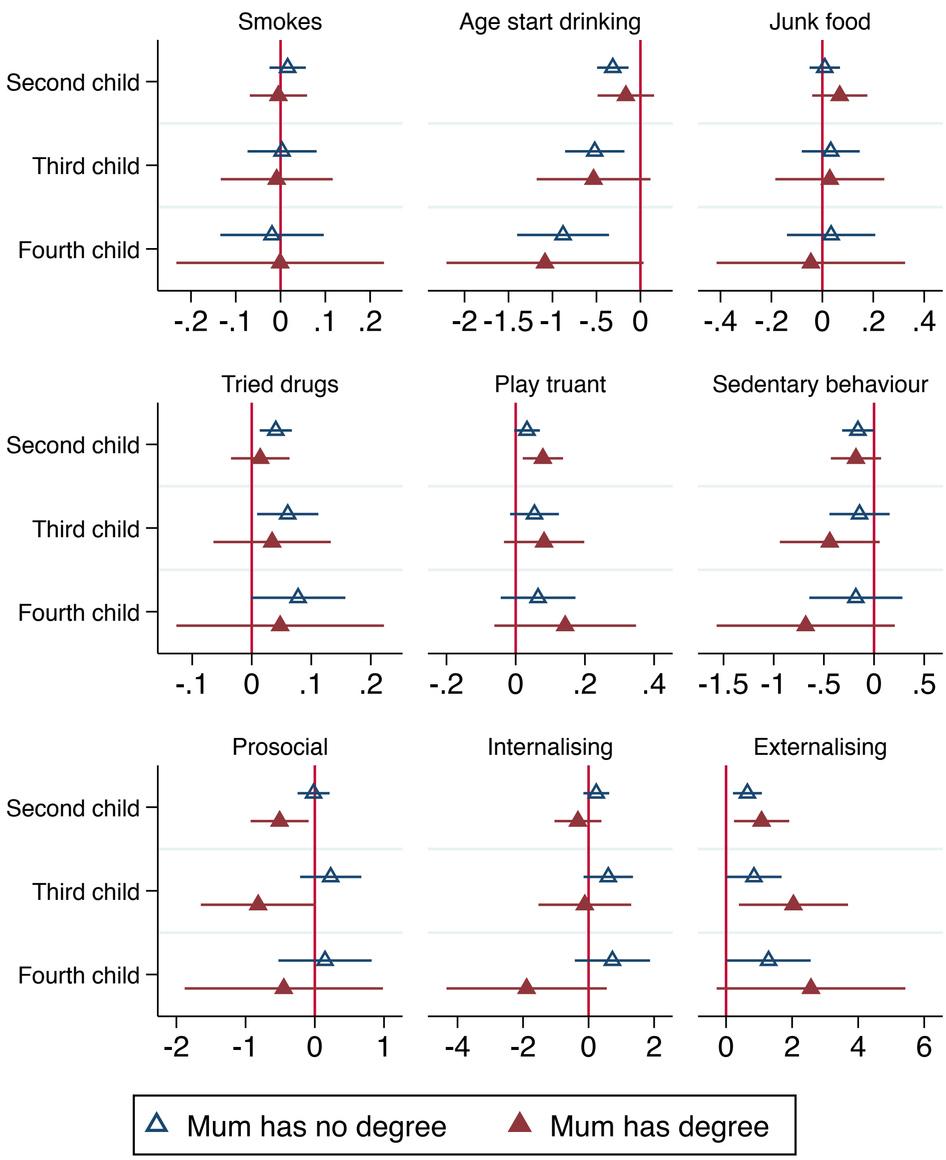


Figure 4.2: Heterogeneity of birth order effects by mother's education. Markers represent point-estimates and lines illustrate respective 95% confidence intervals.

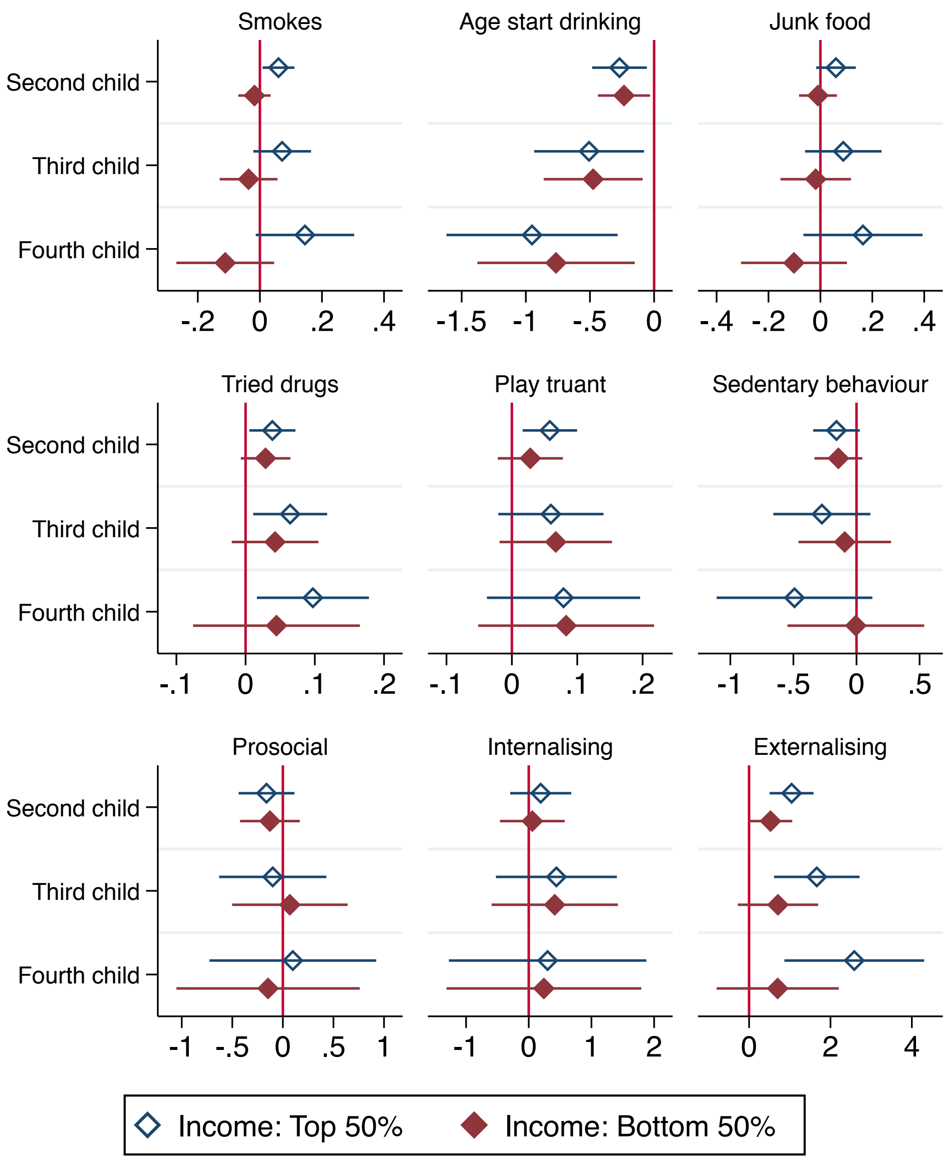


Figure 4.3: Heterogeneity of birth order effects by family income. Markers represent point-estimates and lines illustrate respective 95% confidence intervals.

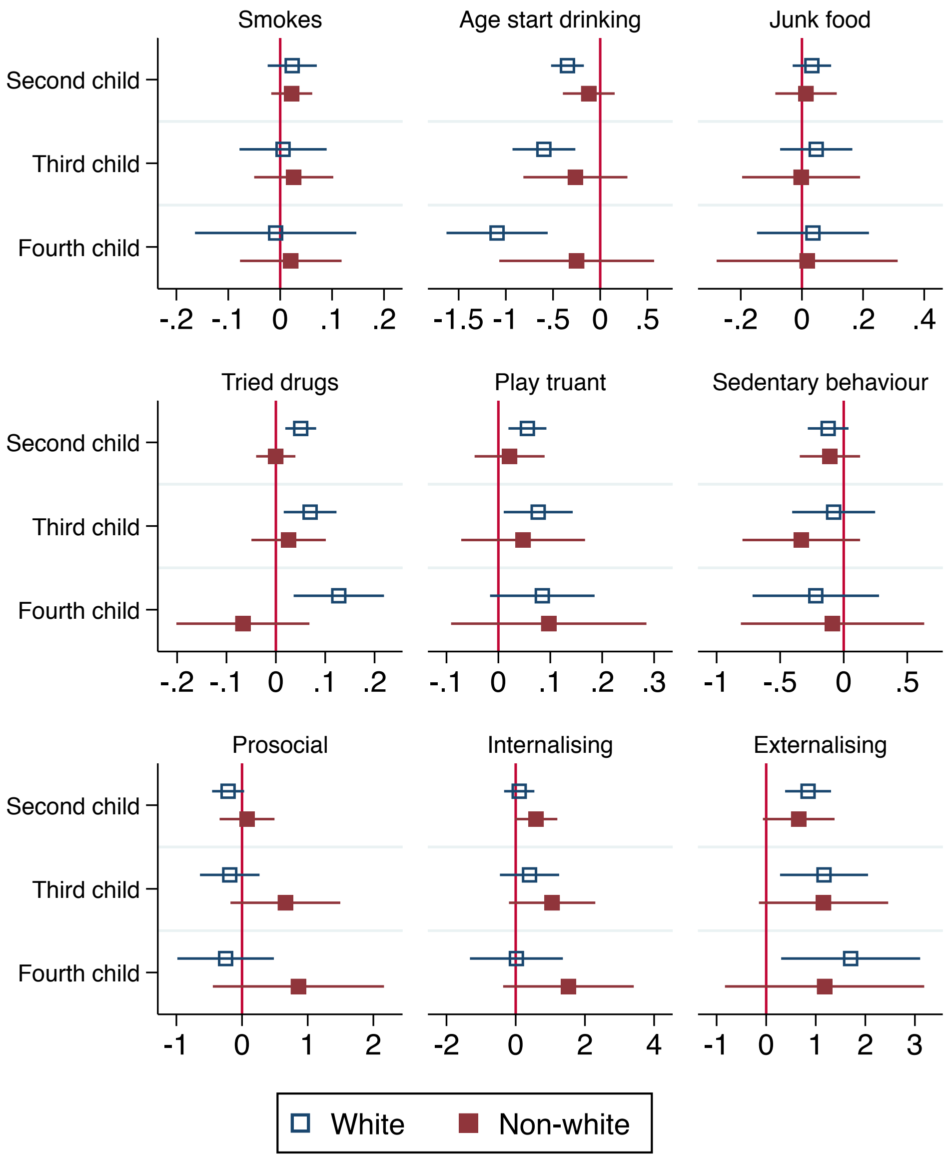


Figure 4.4: Heterogeneity of birth order effects by ethnicity. Markers represent point-estimates and lines illustrate respective 95% confidence intervals.

4.5.3 The contribution of parental investments

The next two sections are concerned with explaining the effect of birth order on risky behaviours and non-cognitive skills by looking at potential channels, starting with parental investments. Differences in parental behaviour by birth order have been observed in the economics literature before, generally finding higher investments for earlier born (Black et al., 2018; Hotz and Pantano, 2015; Lehmann et al., 2018; Monfardini and See, 2016; Price, 2008). By regressing parental investments (instead of child outcomes) on birth order, as detailed in Equation (4.2), I find significant birth order differences in this sample also (see Appendix Table C11). Later-born children perceive significantly lower parental interest in how they do at school, more so when they have older siblings of a different sex. Second, third and fourth born are less likely to say their parents help them with homework compared to first born, regardless of sibship sex composition. Finally, second-born children are less likely to talk to their parents when upset. This effect is driven by children whose older siblings are of the same sex, which may be reconciled with the fact that, when upset, younger girls (boys) may be more likely to talk to their older sister (brother) than to their parents.

As an exploratory analysis to test the role of parental investment, I re-estimate Equation (4.2) with the original outcomes, this time including an index of parental investment P_{ij}, as a regressor. The index is constructed by applying principal component analysis (PCA) to all dimensions of parental investment during adolescence listed in Table 4.2, thus excluding breastfeeding behaviour (details in Appendix Section C.5). The equation is as follows:

$$Y_{ij} = \theta_1 + \sum_{k=2}^{4} \theta_k^{FE} 1(B_{ij} = k) + \theta_6^{FE} P_{ij} + \phi_j + \mathbf{X}'_{ij}\delta^{FE} + \upsilon_{ij}. \qquad (4.4)$$

P_{ij} is potentially endogenous, for example because of reverse causality or because it can be correlated with unobservable parental characteristics, such as preferences, attitudes and beliefs, also affecting the outcomes. Assuming the model is correctly specified, these unobservables are captured by mother fixed effects ϕ_j, and the estimated θ_6^{FE} expresses the extent to which the parental investment index explains away some of the effect of birth order. If so, the estimated θ_k^{FE} in Equation (4.4) will reflect residual differences that are solely due to birth order (e.g. linking back to the placenta hypothesis), although the same caveats mentioned in Section 4.3 apply. Even if the assumption does not hold, analysing the correlation of parental investments with the outcomes can still be informative.

Table 4.6: Birth order and all outcomes, adding parental investments.

	Smokes		**Age started drinking**		**Junk food**	
Second child	0.0161	0.0136	-0.2752**	-0.2640**	0.0267	0.0263
	(0.0237)	(0.0236)	(0.0903)	(0.0898)	(0.0275)	(0.0277)
Third child	0.0139	0.0100	-0.4543**	-0.4442*	0.0369	0.0365
	(0.0406)	(0.0404)	(0.1762)	(0.1759)	(0.0520)	(0.0521)
Fourth child	-0.0092	-0.0158	-0.8174**	-0.8028**	0.0305	0.0299
	(0.0593)	(0.0589)	(0.2802)	(0.2792)	(0.0803)	(0.0806)
Female	0.0023	0.0009	0.1171+	0.1252+	-0.0455*	-0.0457*
	(0.0156)	(0.0156)	(0.0711)	(0.0714)	(0.0196)	(0.0196)
Par. investments		-0.1032*		0.3545+		-0.0119
		(0.0507)		(0.1906)		(0.0581)
Fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3282	3282	5144	5144	5993	5993
	Tried drugs		**Play truant**		**Sedentary behaviour**	
Second child	0.0333*	0.0325*	0.0324+	0.0268	-0.1556*	-0.1575*
	(0.0141)	(0.0142)	(0.0184)	(0.0182)	(0.0766)	(0.0768)
Third child	0.0483*	0.0479+	0.0373	0.0310	-0.2502+	-0.2519+
	(0.0245)	(0.0245)	(0.0331)	(0.0329)	(0.1514)	(0.1516)
Fourth child	0.0381	0.0377	0.0579	0.0476	-0.2878	-0.2912
	(0.0434)	(0.0433)	(0.0500)	(0.0499)	(0.2301)	(0.2300)
Female	-0.0160+	-0.0168+	-0.0025	-0.0065	0.4456***	0.4440***
	(0.0097)	(0.0097)	(0.0129)	(0.0127)	(0.0568)	(0.0566)
Par. investments		-0.0366		-0.1760***		-0.0675
		(0.0328)		(0.0436)		(0.1660)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4368	4368	6005	6005	5689	5689
	SDQ Prosocial		**SDQ Interalising**		**SDQ Externalising**	
Second child	-0.1334	-0.0768	0.1647	0.0742	0.8266***	0.6799***
	(0.1049)	(0.1038)	(0.1834)	(0.1826)	(0.1979)	(0.1922)
Third child	-0.0287	0.0315	0.5512	0.4558	1.2468***	1.0923**
	(0.2030)	(0.1997)	(0.3621)	(0.3589)	(0.3743)	(0.3608)
Fourth child	-0.0838	0.0165	0.5563	0.3960	1.8085**	1.5493**
	(0.3180)	(0.3113)	(0.5748)	(0.5632)	(0.5784)	(0.5599)
Female	0.9319***	0.9732***	1.0899***	1.0218***	-1.0936***	-1.2051***
	(0.0769)	(0.0748)	(0.1269)	(0.1255)	(0.1430)	(0.1359)
Par. investments		1.7904***		-2.9010***		-4.7493***
		(0.2511)		(0.3905)		(0.4362)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6001	6001	5998	5998	5998	5998

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

Table 4.6 shows estimation results for Equation (4.4), for all outcomes. For each outcome, the first column presents estimates of the main fixed-effect specification, while the second column includes the parental interest dummy[17]. Parental investments are significantly associated with three out of the six behavioural outcomes. Higher parental investments are linked to a lower probability that the child smokes or plays truant, and to a higher age of first alcoholic drink. Additionally, including parental investments changes the magnitudes of the coefficients associated to birth order: the coefficients for second-born children are reduced by 4% and 6% respectively for age of first drink and play truant. Given these small magnitudes, these results suggests that parental investments may not be the only channel to understand birth order differences in unhealthy and risky behaviours. Parental investments are also strongly associated with higher non-cognitive skills in all three dimensions. Moreover, differences in parental interest by birth order account for about 20% of the effect of birth order on the SDQ externalising score[18].

4.5.4 Sibling interactions

Next, I assess the role of sibling interactions as a mechanism to explain birth order differences in the outcomes. The rationale for this analysis is to address the arguments advanced by psychology theory, discussed in Section 4.2.2. When engaging in a risky behaviour, older siblings could be setting an example for younger ones, facilitating onset of the same behaviour, compared to other children of the same age who do not have older siblings behaving in such ways. On the other hand, as they acquire more information about the potential negative consequences of risky behaviours, younger siblings could be deterred from engaging in such actions. I thus test, on the one hand, for patterns of imitation in younger siblings' behaviours, and on the other, for any effect on younger siblings' awareness of the dangers associated to risky behaviours.

I first look at whether younger siblings' behavioural outcomes are associated with a binary indicator that equals one if the eldest sibling observed in the family engages in a risky behaviour, and 0 otherwise[19]. I carry out the analysis for smoking, drinking and trying drugs using an OLS specification, controlling for all covariates. I cannot directly introduce mother fixed effects, since the explanatory variable referring to the eldest sibling's behaviour does not vary across children

[17]Sample sizes are smaller than in Tables 4.4-4.5 due to missing data in the parental investment variable.

[18]In alternative specifications I substitute P_i with each parental investment in turn, obtaining similar results. The main difference is that the magnitude of birth order coefficients is reduced less than when using the investment index.

[19]Substituting this indicator with one for any older sibling gives qualitatively similar results.

Table 4.7: Eldest sibling's behaviour, risky behaviours and beliefs about risky behaviours

	Younger sib.: Risky behaviours					
	Smokes		**Age start drinking**		**Drugs**	
Eldest smokes	0.5507***	-0.0008				
	(0.0283)	(0.0031)				
Eldest drinks			-0.7366***	-0.0559***		
			(0.0560)	(0.0092)		
Eldest tried drugs					0.5502***	0.0193***
					(0.0269)	(0.0048)
Predicted FE		0.9998***		0.9844***		0.9753***
		(0.0032)		(0.0034)		(0.0071)
Estim. FE term	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3223	3223	3860	3860	3135	3135
	Younger sib.: Beliefs about risky behaviours					
	Heavy smoking		**Heavy drinking**		**Cannabis regularly**	
Eldest smokes	-0.0531	0.0012				
	(0.0464)	(0.0020)				
Eldest drinks			-0.1179***	-0.0043+		
			(0.0204)	(0.0024)		
Eldest tried drugs					-0.3628***	-0.0108*
					(0.0425)	(0.0046)
Predicted FE		0.9996***		0.9987***		0.9993***
		(0.0017)		(0.0024)		(0.0028)
Estim. FE term	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4251	4251	3270	3270	2853	2853

Standard errors clustered at mother level in parentheses. Beliefs are binary variables.
+ $p < 0.01$ * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

from the same mother. Instead, I include as a regressor the fixed-effect term predicted in the main regressions, shown in Tables 4.4 and 4.5. The idea is that the predicted fixed-effect term should capture enough of the unobservable family-specific variation to make the coefficient of interest closer to the true effect of sibling interaction[20]. First-born children are excluded from this analysis, since they do not have a sibling they can look up to. The top panel of Table 4.7 shows that each of the three indicators for eldest sibling's behaviour is significantly associated with the same behaviour in younger siblings. Without accounting for unobservables, having an eldest sibling who engages in the corresponding behaviour is associated with a 55% increase in the probabilities of smoking and trying drugs, and to a reduction in age of first drink of 0.7 years (approximately 8 months). The association is smaller once the predicted fixed-effect term is included. Partially controlling for unobservable family-specific traits makes the association with the indicator for eldest sibling smoking null. The association between the indicator for eldest sibling drinking and younger sibling's drinking age now declines to -0.06 years (approximately 0.7 months earlier). Finally, if

Table 4.8: Birth order and beliefs about risky behaviours

	Beliefs about risky behaviours					
	Heavy smoking		**Heavy drinking**		**Cannabis regularly**	
Second child	-0.0105	-0.0249	-0.0052	-0.0118	-0.0381**	-0.0757**
	(0.0143)	(0.0285)	(0.0161)	(0.0314)	(0.0140)	(0.0291)
Third child	0.0005	-0.0098	-0.0061	-0.0276	-0.0781***	-0.1270*
	(0.0223)	(0.0580)	(0.0245)	(0.0601)	(0.0226)	(0.0575)
Fourth child	-0.0562	-0.0824	-0.0912*	-0.1950+	-0.0805*	-0.1126
	(0.0403)	(0.0961)	(0.0442)	(0.1025)	(0.0402)	(0.0923)
Fixed effects	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4996	4996	4904	4904	4732	4732

Standard errors clustered at mother level in parentheses. Beliefs are binary variables.
+ $p < 0.01$ * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

the eldest sibling has tried drugs, now only a small 2% increase in probability of the younger one taking drugs is observed. The fixed-effect term has a large and significant coefficient, indicating that, although there is evidence of sibling imitation patterns, family-level traits still explain a large portion of the variation in younger siblings' behaviour.

Secondly, I check whether the eldest sibling's behaviour affects younger siblings' awareness about the risk attached to heavy smoking, heavy alcohol drinking and heavy drug use. I exploit survey questions in the Young Person element of Understanding Society, where children are asked to rate the riskiness of several behaviours on a 0-4 scale, from "No risk" to "Great risk" (more details on these variables in Appendix Section C.6). The belief outcome variables are recoded so they are dummy variables, taking the value one if the child rates the dangerous behaviour as carrying great risk, and zero otherwise. Results in the bottom panel of Table 4.7 show that observing older siblings drinking or taking drugs reduces younger siblings' perception of the risk attached to heavy drinking and regular cannabis use respectively. No effect is observed on beliefs related to smoking. Once the fixed-effect term is included, the coefficient for eldest sibling drinking is a negligible -0.004, while the coefficient for eldest sibling taking drugs decreases to 0.01. Both terms are still significant, but their importance decreases considerably, especially compared to family-level characteristics, which again appear to have a large and significant relationship with children's beliefs about the riskiness of these behaviours.

Finally, I check if these beliefs present differences by birth order, to further understand whether they are part of the birth order effects story. I regress the belief variables on the birth order indicator, estimating a model similar to Equation (4.2) again. Results in Table 4.8 show that birth order differences are only

[20] I am grateful to Thomas Cornelissen for this suggestion.

significant for the risk attached to daily cannabis smoking, with later born perceiving it as less risky, compared to first born. The main implication of these findings is that part of the birth order effect on drug-taking is to be attributed to imitation patterns by younger siblings, who are less informed or less averse when it comes to the risk involved in trying drugs. Information about the dangers of heavy drinking is also marginally affected by older siblings' behaviour, although this may not be reflected in differences in beliefs among siblings.

4.5.5 Ruling out optimal stopping

As discussed, the fixed-effect strategy takes care of potential unobserved confounders that are constant across siblings. However, there is one source of endogeneity of fertility decisions that is not ruled out by fixed effects. Parents may initially have a preference for a given number of children, but they may update their preference after they obtain information on each new child. This is especially likely if a child takes up more resources than expected, for example because of problematic behaviour or sickness (Rosenzweig and Wolpin, 1988). If the event of a problematic child stops parents from having more children, I may never observe later children in such families. This would lead to excluding the family from the sample or to a spurious negative effect of being the youngest child[21]. On the other hand, if a good child increases the likelihood of having more children, say because parents enjoy parenthood more than they expected, I may find a spurious negative effect of being born later, only because such child would be "above average quality", while subsequent ones would be of "average quality"[22].

To the extent that these are rare events, or that birth spacing is small and therefore parents do not observe outcomes for earlier born before they have another child, this may not seriously bias treatment effect. Yet, it is still desirable to check if there is evidence of "optimal stopping behaviour" by parents, in order to increase robustness of the study[23]. I thus estimate a linear probability model, regressing the probability of having younger siblings, $Pr(YS_i)$, on a child's SDQ difficulty and prosocial scores, including controls for background and birth order of the child. The SDQ difficulty score is preferred here to the in-

[21]The family would be excluded from my sample if the problematic child was the first born, as it would be a single-child family. Alternatively, it would bias the estimator negatively if the problematic child was any other birth order.

[22]A similar argument is made by Black et al. (2018).

[23]Three studies in the relevant literature have tested for this type of behaviour, two of them by regressing the probability of having another child on first-born children's outcomes (Björkegren and Svaleryd, 2017; Black et al., 2018; Pavan, 2016). Black et al. (2018) adopt an original approach by simulating a hypothetical second child for all families with one child, assuming the extreme scenario that all single-child families had 'optimally stopped' having children, and showing that their estimates are robust to this hypothesis.

ternalising/externalising categorisation, to better capture problematic behaviour.

$$Pr(YS_{ij}) = \eta_0 + \eta_1^{FE}\text{SDQDIFF}_{ij} + \eta_2^{FE}\text{SDQPRO}_{ij} + \phi_j + \mathbf{X}'_{ij}\delta^{FE} + \upsilon''_{ij}. \quad (4.5)$$

If the estimated parameters η_1^{FE} and η_2^{FE} are significantly different from zero, then this could be taken as evidence that earlier children's outcomes influence parents' fertility decisions, biasing birth order effect estimates. Vector $\mathbf{X}_{ij}$ in Equation (4.5) also contains the birth order dummies B_{ij}, in order to account for the fact that the probability of having younger siblings changes as family size increases. The choice of SDQ scores over other outcomes to test for optimal stopping is due to the fact that, presumably, parents would be able to notice their child's problematic behaviour early on, which could then in turn affect their subsequent fertility decisions. While SDQ measures are expected to be highly correlated to early signs of problematic behaviours, other risky behaviours are initiated later on, when parents' fertility is often completed.

Table 4.9 shows linear probability models for optimal stopping behaviour, pooling together 2-, 3- and 4-child families[24]. The first three columns present OLS estimates, while the last three show fixed-effect results, progressively adding controls in both cases. The two variables of interest, SDQ difficulty and prosocial scores, are both negatively and significantly associated with the probability of having younger siblings, which indicates a degree of endogeneity of fertility decisions. This is only partly a surprise: high difficulty scores may be expected to decrease the probability of having more children. Conversely, if a child is very prosocial, the expected effect on parental fertility decisions may be positive. When controlling for SDQ scores only, for a 1-point increase in the child's difficulty score, there is only a 0.2% decrease in the probability that parents will have more children. For a 1-point increase in the prosocial score, perhaps surprisingly, the probability of another child decreases by 1%. The small coefficients indicate that the bias introduced by optimal stopping is potentially quite small. Reassuringly however, once I include mother fixed effects and controls as in my preferred specification, the coefficients on the SDQ scores are not significant any more, indicating that this strategy is successful in dealing with confounders and any residual endogeneity of parental fertility decisions.

[24] One-child families are not included in the analysis. Any optimal stopping behaviour on their part would not bias the estimates, given that the main analysis is concerned with families who have between two and four children.

Table 4.9: Testing for optimal stopping behaviour. Probability of having any younger sibling and older children's SDQ score.

	Dep. variable: Probability of having any younger siblings					
SDQ difficulty	-0.0019+	-0.0016*	-0.0013*	-0.0078**	-0.0012	-0.0015
	(0.0011)	(0.0007)	(0.0006)	(0.0027)	(0.0016)	(0.0015)
SDQ prosocial	-0.0123***	-0.0055**	-0.0042*	-0.0290***	-0.0063	-0.0072
	(0.0034)	(0.0020)	(0.0021)	(0.0076)	(0.0044)	(0.0046)
Second child		-0.5690***	-0.5694***		-0.5800***	-0.4741***
		(0.0083)	(0.0101)		(0.0128)	(0.0210)
Third child		-0.9683***	-0.9698***		-1.1162***	-0.8970***
		(0.0121)	(0.0149)		(0.0240)	(0.0425)
Fourth child		-1.4060***	-1.4112***		-1.8760***	-1.5547***
		(0.0205)	(0.0187)		(0.0314)	(0.0595)
Female			0.0002			0.0178
			(0.0072)			(0.0151)
Constant	0.6324***	0.0660**	0.2062***	0.8228***	1.0254***	0.5252+
	(0.0317)	(0.0220)	(0.0505)	(0.0712)	(0.0406)	(0.2943)
Fixed effects	No	No	No	Yes	Yes	Yes
B.O. + sibship	No	Yes	Yes	No	Yes	Yes
Controls	No	No	Yes	No	No	Yes
Observations	6864	6864	6864	6864	6864	6864

+ $p < 0.01$ * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Standard errors clustered at mother level in parentheses. SDQ difficulty, ranging from 0 to 40, is the sum of the internalising and externalising scores.

4.6 Discussion

4.6.1 Putting results into context

I find a robust link between birth order and most behaviours surveyed, showing that later-born children are more likely to start drinking alcohol earlier, skip school, try drugs and do more exercise. This result largely confirms the associations observed by Argys et al. (2006) in US adolescents, but has the advantage of a robust fixed-effect strategy, which minimises the possibility of bias due to family-level unobservable characteristics. Both Argys et al. (2006) and Black et al. (2016) found higher smoking rates for later born while, in the present context, birth order differences in smoking in adolescence are largely eliminated by accounting for family characteristics. Something similar can be said for junk food consumption, which makes sense, considering that diets are often parental choices and thus they may not vary across siblings.

Secondly, the study provides strong evidence that non-cognitive skills are lower for later-born children compared to earlier born, with the former more likely to have high externalising scores, predictive of aggressiveness and anger. In some specifications, later born also display significantly lower prosocial scores, increasing in non-cognitive skills. Black et al. (2018) find a similar result for a different measure of non-cognitive skills in Swedish 18-year-olds, while Lehmann et al. (2018) find no evidence of birth order effects on an index of behavioural problems, temperament and general self-perception in US children up to age 14. However, when analysing birth order effects separately by age group, Lehmann et al. (2018) do find some evidence of a negative effect of higher birth order in the later age group, 11-14. Together with the present evidence and Black et al. (2018), this may suggest that differences in non-cognitive skills arise in adolescence, and persist into adulthood. Observed differences in externalising score by birth order could translate into differences in risky behaviours by birth order, designating non-cognitive skills as a primary explanatory channel for birth order effects. Moreover, childhood internalising and externalising scores have been shown to be highly predictive of adult educational and labour market outcomes, with high externalising scores linked to higher probability of dropping out of school, reduced enrolment in higher education and higher sickness absence (Evensen et al., 2016; Narusyte et al., 2017).

In terms of magnitude, birth order effects are relatively large. Taking mean values for reference, being born second represents a 100% increase in the average probability of trying drugs compared to the average for first born (from 3% to 6%), and a 40% increase in the average probability of playing truant at school (from about 10% to 14%). Ceteris paribus, average externalising scores increase by 13% for second born, 20% for third born and 28% for fourth born, compared to first born. Moreover, significant gender differences in risk-taking and non-cognitive skills are widely documented in the literature, and can therefore serve as a benchmark against which to assess the magnitude of birth order effects (Byrnes et al., 1999; Charness and Gneezy, 2012; Maguire et al., 2016; Reid et al., 2000). For instance, the difference in the average probability of trying drugs is twice as large between first and second born than between boys and girls. Conversely, the difference in average externalising score between first and second born is half the difference between boys and girls.

I further find an effect of sibship sex composition on some outcomes. Effects are generally larger for children with at least one older sibling of a different sex: this is true for prosocial and externalising scores, for trying drugs and for skipping school. Only for age of first drink the coefficient is larger for children with older sibling of the same sex. The literature offers mixed findings on whether sex composition is important. While Argys et al. (2006) find no association between sex compositon and behaviour, other studies have found better outcomes for girls whose older siblings are boys (Butcher and Case, 1994; Tenikué and Tequame, 2017). Overall, results show that the main driver of birth order effects is being the first child versus being born later. Anecdotal evidence in the popular culture about rebellious middle children would appear not to be confirmed by the data.

Several studies have detected some heterogeneity in birth order effects with respect to individual background and child sex. Similarly to existing literature, I find that birth order effects are slightly more prominent for boys (Argys et al., 2006; Black et al., 2018; Rees et al., 2008). The finding that effects may be stronger for families of higher socio-economic status is corroborated by some of the literature (Barclay et al., 2017), but different results have also been highlighted in other contexts (Björkegren and Svaleryd, 2017; Bonesrønning and Massih, 2011).

4.6.2 Sources of birth order effects

The study also provides some pointers to explain why higher birth order children engage more in risky behaviours. The finding that parental investment measures differ by birth order is consistent with several studies (Black et al., 2018; Hotz and Pantano, 2015; Lehmann et al., 2018; Mechoulan and Wolff, 2015; Pavan, 2016; Price, 2008)[25]. Younger children talk to their parents less when they have siblings of the same sex, maybe because girls (boys) prefer talking to their older sister (brother) about things that matter to them. This suggests that sibling sex composition could account for parent-child relationship and child behaviour, an original finding in this literature. Both individually and combined into an index, parental investments are positively associated with non-cognitive skills, and they account for roughly 20% of the birth order effect. Parents may actively invest more in first born, and progressively less for each additional child, producing differences in non-cognitive skills, consistently with the resource dilution hypothesis advanced by Blake (1981). Parental investments were only directly linked to some of the risk-taking behaviours in adolescence, and therefore other mechanisms should be explored to further understand birth order effects on these outcomes.

Secondly, to the best of my knowledge, this is the first study that tries to directly assess sibling interactions as a mechanism to explain birth order effects. The descriptive results presented show that children whose eldest sibling drinks regularly or has tried drugs are likely to engage in the same behaviour, although it is acknowledged that these results could be biased by family-level confounders. Interestingly, the findings are in line with a study by Altonji et al. (2017) on US adolescents, finding a small but significant causal effect of older siblings' behaviour

[25]Other studies found significant differences in breastfeeding by birth order (Black et al., 2016; Buckles and Kolka, 2014; Lehmann et al., 2018). I do not find the same here, but this is potentially linked to the UK having a less rooted tradition of breastfeeding than other European countries and the US (World Health Organisation, 2019).

on younger siblings' drinking behaviour and likelihood of marijuana use. Another possible consequence of being exposed to older siblings' behaviour are changes in younger children's risk awareness. Observing older siblings engage in risky behaviours is found to lower awareness about the risks attached to drinking and consuming cannabis regularly, with the latter being perceived as significantly less risky by later-born children, compared to first born.

Explaining birth order effects via parental investments and sibling interactions can inform policies aimed at improving early childhood environment, to foster individual development and prevent risky behaviours in adolescence. Evidence on the dilution of parental resources can help re-assess baby bonus policies, for instance by increasing publicly funded per capita resources for higher order children. Since birth order differences are observed within the family, for children sharing the same household and resources, even small incentives for parents could help equalise the cost of these investments across children. Examples of useful measures to decrease the cost of parental investments in later-born adolescent children are tax credits on specific goods for larger families or family-friendly policies at the workplace. On the other hand, the significance of peer interactions supports the implementation of peer programmes where older children directly encourage healthier habits in younger children, another potentially effective tool that has found increasing popularity in schools (Campbell et al., 2008; Foley et al., 2017). Examples of such programmes can be found in school and community settings in different countries, and they have potential to also generate positive spill-over effects on peer educators (Strange, 2002).

4.7 Conclusion

Using data from a panel of UK households, this paper finds evidence that later birth order is linked to a higher probability of engaging in risky behaviours and to lower non-cognitive skills in adolescence. I have dealt with two of the main issues faced in this literature, separating the effect of birth order from that of family size and reducing the endogeneity of fertility decisions due to unobservable family characteristics.

The exploration of sibling interactions in the birth order literature is original, since parental investments have usually been the main channel researched to understand birth order effects. Studies on sibling dynamics that try to isolate the causal effect of the interaction are rare and could be the subject of fruitful future research. For instance, it would be interesting to identify more at-risk groups, and to explore differences in timing of risky behaviour uptake, as such behaviours can be riskier, the earlier a child engages in them. Another avenue for future research is to explore whether effects on risk-taking by birth order last over time, and what implications they have for future lives. Bertoni and Brunello (2016) analyse lifetime earnings, and find that birth order effects are only observed in the first ten years in the labour market, and dissipate from then onwards. Findings by Black et al. (2016) and Black et al. (2018) on adult non-cognitive skills and health behaviours suggest these effects in such domains are long-lasting.

The decisions people make about their education, job, relationships, their health and everyday lives are often affected by how much risk they are prepared to take. Findings of this study are important because adolescence is a time at which individuals are particularly prone to adopting risky behaviours that can then affect them in their adult life. The result that higher birth order is linked to higher risk-taking in adolescence can offer an early insight on future decisions, as well as complementing previous literature showing better educational, labour market and adult health outcomes for first born compared to later born. Findings suggest that within-family variation is an important aspect to consider when devising policies aimed at reducing risky adolescent behaviour.

More widely, the paper sheds light on the early origins of risk-taking behaviours, and how they are linked to the household environment and to the behaviours of other agents in the household, in a setting where the effect of the background is robustly controlled for and the genetic mix is randomly assigned.

Chapter 5

Conclusion

This thesis explores dimensions of schooling and family environment as determinants of human capital, mainly focusing on long-term health and well-being outcomes. In the choice of topics covered, priority is given to areas that can be influenced by policy, and thus effectively changed, with an eye to their relevance for current affairs. Chapters 2 and 3 analyse the consequences of selective schooling in England, exploring its average effect and the effect for directly affected pupils respectively. Chapter 4 looks at the role of birth order in explaining variation in risky behaviours and non-cognitive skills in adolescence, an understudied but crucial period in the human capital accumulation process.

Chapter 2 shows that the average long-term impact of going to school in a selective system, compared to a mixed-ability system in the 1970s England, differs by prior cognitive ability. In the high cognitive ability sample, who can access grammar school if exposed to selection, selective schooling increases the average aspirations towards academic achievement, adult wages and the probability of employment. However, results also suggest that selective schooling in this sample could lower average adult life satisfaction. In the lower cognitive ability sample, who experience lower average peer ability and school quality when exposed to selection, selective schooling lowers school aspirations, while it marginally raises wages and self-efficacy. This result may be due to the practical nature of the secondary modern curriculum at the time, more geared towards specific types of professions. The chapter also shows that selective schooling is not directly linked to the majority of the long-term health and well-being outcomes surveyed, which are instead associated with childhood cognitive and non-cognitive abilities..

Chapter 3 answers the question of the impact of attending a selective secondary school within the selective system, for the restricted group of pupils who are at the margin of being admitted. Building on the findings of Chapter 2, Chapter 3 focuses on a narrower range of cognitive ability, looking at individuals most likely to be affected by an expansion in grammar school places. By exploiting the data in an innovative way, a quasi-experimental framework can be implemented, using for the first time an RDD methodology to answer this question for several regions in England. Results show that even when focusing on a restricted section of the ability distribution, grammar school is only a significant predictor of higher academic attainment for the marginal admitted student. Strikingly, this effect is only found for pupils displaying high socio-economic status or high mother interest in their education. Other health and well-being outcomes are not affected. Thus, although educational attainment is generally a determinant of labour market success and well-being later in life, for the sample in question, the difference in A-levels (and potentially probability of obtaining a university degree) produced by grammar attendance was not sufficient to predict significant differences in other long-term human capital outcomes.

Findings of both chapters are based on historical data on a generation that faced very different circumstances to those experienced by young generations nowadays. Drawing the implications of these findings for present policy requires several adjustments, starting from considering how the features of the English school system have changed since the 1970s. First, there are now more types of publicly funded schools, including free schools, academies and faith schools, widening the set of options and therefore possibly changing how the most able pupils are distributed across school types. Second, the emphasis on A-levels and higher education is now more widespread across different school types, with 28.6% of school-leavers going straight to university in 2017 compared to 5.57% in 1974 (House of Commons, 1976; UK Department for Education, 2019a)[1]. Third, vocational subjects have long stopped being the focus of the secondary modern curriculum, which in line with the national curriculum offers all standard 'core' and 'foundation' academic subjects, thus lowering barriers to further study for their pupils compared to the past (UK Department for Education, 2014).

In its first 2018/2019 round, the Selective Schools Expansion Fund supported 16 expansions projects, for a total of £49.3 million spent towards the creation of 2,700 new grammar school places. Applications for the second £50 million round are under review at the time of writing, as part of a total projected £200 million expenditure over four years (UK Department for Education, 2019b). Based on the findings in Chapters 2 and 3, if the same circumstances applied to the current generation of pupils, then for every £50 million spent, just under 3,000 pupils would in principle benefit from better education outcomes, with the caveats mentioned. Given the differences mentioned above, we acknowledge that the consequences of the grammar system in the past cannot be transposed straightforwardly to cur-

[1]In 2017/2018, the percentage of school-leavers registering for a first degree by age 21 rises to 48%, and to 50.2% by age 30.

rent policy. For instance, given the current emphasis on higher education across the whole public secondary school system, it may be that the marginally admitted pupils would go to university anyway, and that grammar attendance would improve their educational opportunities by affecting the rank of the institution or subject studied instead (Burgess et al., 2017). Further, according to findings in Chapter 3, the positive effect on educational outcomes could be conditional on having a favourable family background. On the other hand, the positive average effects of both school types in the selective system on wages and employment found in Chapter 2 may have been nullified over time, as literature using more recent datasets would suggest (Burgess et al., 2019). One hypothesis is that this is due the progressive migration of non-selective schools in selective areas towards a more standard academic education, and the decline of the value associated to vocational qualifications, now largely left to colleges and company-based apprenticeships (Wolf, 2011).

Additionally, the current measure is expected to affect the intake of publicly funded non-selective schools in the same areas as grammar schools, by lowering the average ability of their pupils, and possibly causing some migration of the most able teachers to newly expanded grammar schools (Allen, 2016a). According to a study by Allen (2016b), every year at least 281 public schools saw their intake affected by neighbouring grammar schools prior to the expansion policy, and the number is likely to rise with more selective places being created. Based on anecdotal evidence from the past, the move may also expose a higher number of pupils to the disappointment of not passing an exam perceived to determine their life chances (Skipper and Douglas, 2016; The Guardian, 2015).

Whether it is preferable to invest public resources in marginally improving the chances of the targeted high-ability individuals or whether other groups should be prioritised is largely a normative question. It should be noted that, together with the Selective School Expansion Fund, other measures promoted by the government 2016 'Schools that work for everyone' strategy included encouraging selective schools to support non-selective ones and to increase access for disadvantaged pupils (UK Department for Education, 2018). However, it remains an open question whether these objectives are more efficiently achieved by funding selective schools, rather than giving funds directly to non-selective schools or those operating in disadvantaged areas. The findings in this thesis suggest that the gains associated to grammar school for a 60-year-old generation were only marginal, conditional on their background, and mainly concerned further educational opportunities and marginally higher wages, while, when isolated from other influential factors, type of school did not affect other dimensions of health and human capital.

Finally, results point towards childhood cognitive and non-cognitive skills as a promising avenue for improving human capital. Funding interventions to improve these earlier outcomes could be a way to distribute public money more in line with enhancing equality of opportunity early on. Evidence from developmental research has found early childhood programmes to be particularly effective in improving cognitive and non-cognitive outcomes, and the attention for the first five years of life has risen in policy making in recent years (Campbell et al., 2014; Currie and Almond, 2011; Heckman et al., 2013; Kautz et al., 2014). As already mentioned, all that can be said at present relies on data collected in the past. To assess the actual effect of the current grammar school expansion, researchers will have to wait for data on the generation exposed to it. Present efforts to link administrative data from the educational sphere to health and social security data in the UK will prove fruitful in answering this question.

Chapter 4 moves to family environment, aiming to explore how earlier circumstances could affect individual non-cognitive skills and risky behaviours in adolescence, both important determinants of adult human capital and health. Order of birth of children in the family was shown to be a significant predictor of adolescent outcomes, lowering average non-cognitive skills and increasing the probability of engaging in some risky behaviours, such as early age alcohol drinking, drug taking and skipping school. The largest gap was generally found between first-born children and their younger siblings. The effect is larger for boys and for higher socio-economic status families, which is speculatively attributed to larger constraints to parents' time with their children in such families. Overall, parental behaviour was shown to vary by birth order, and to explain away part of the non-cognitive skill advantage of firstborn. The analysis also presented preliminary evidence that, as hypothesised, children imitate their older siblings' risky behaviours, in reference to smoking, drinking alcohol and trying drugs. An explanation for this effect is the lower perception of risk attached to several risky behaviours by children with older siblings.

Finding evidence of human capital differences within the same family can inform better family policies. Contrarily to cross-family differences, which are also important, within-family differences may be easier to eliminate with the necessary support or right incentives for families, thus delivering effective improvements to children's human capital. In line with the results of Chapter 4 and other related literature, there might be a case for parental leave or baby bonus policies to have specific features targeting larger families, to reduce parental resource dilution as more children are born, which appears to be an issue. Given that the differences in parental behaviours highlighted in the data concern investments during adolescence, policies targeting this period could be particularly effective. Example measures are more flexible work arrangements for parents and progressively larger child benefits or tax credits for higher numbers of children. Often, first time parents benefit from extra public resources, but this research makes a case for more resources to be diverted to later births too.

A further important finding of the paper concerns the significance of sibling interactions for risky behaviours. Future research is needed to further uncover the mechanisms by which imitation patterns occur, and to identify more at risk groups. For instance, it would be interesting to investigate whether the timing of uptake of risky behaviours coincides between siblings, or whether siblings are more likely to imitate an older sibling of the same versus opposite sex. Clarifying the key times and mechanisms for peers' influence during adolescence would strengthen support for school- or community-based peer education programmes to prevent adolescent risky behaviours. Effective examples of such programmes include ASSIST, a UK programme for smoking reduction, and SALSA, an Australian programme for improving diet and physical activity, both school-based and delivered by older students to younger cohorts (Campbell et al., 2008; Foley et al., 2017). Moreover, these programmes have been shown to have a positive impact on peer educators themselves, who are also generally adolescents (Strange, 2002). Studies on the cost-effectiveness and national scalability of such interventions would be useful to inform policy practice. Lastly, the paper does not analyse the significance of sibling interactions for non-cognitive skill formation, and this could be the subject of productive future research work.

In an ideal world, all policy questions would be evaluated in a randomised control trial (RCT) that allows for all the specificities of each case. However, due to the nature of several social issues and to the high monetary and time costs, RCTs are often infeasible. It is still useful to do research on these issues, and provide answers, which can always be improved by the occurrence of new data, methods and resources. This thesis has provided evidence on returns to schooling and family environment, two areas that are traditionally affected by the standard endogeneity problem, complicating the estimation of their effect on individuals' outcomes. Its findings help make a case for specific policies for the improvement of human capital and health, while at the same time spurring further questions for future research. One of them concerns the significance of educational policy for the improvement of health and well-being, given that an important aspect of school quality has been shown not to affect these dimensions. Both quality and quantity of schooling were thought to affect health productivity in the seminal model by Grossman (1972), but the empirical literature does not always corroborate this theory (Clark and Royer, 2013; Galama et al., 2018). Instead, family background consistently accounts for differences in health, well-being and human capital more in general. Identifying the different mechanisms within the family environment that influence future outcomes to inform more cost-effective interventions constitutes the challenge of future human capital research.

Appendix A

Appendix to Chapter 2

A.1 Appendix figures

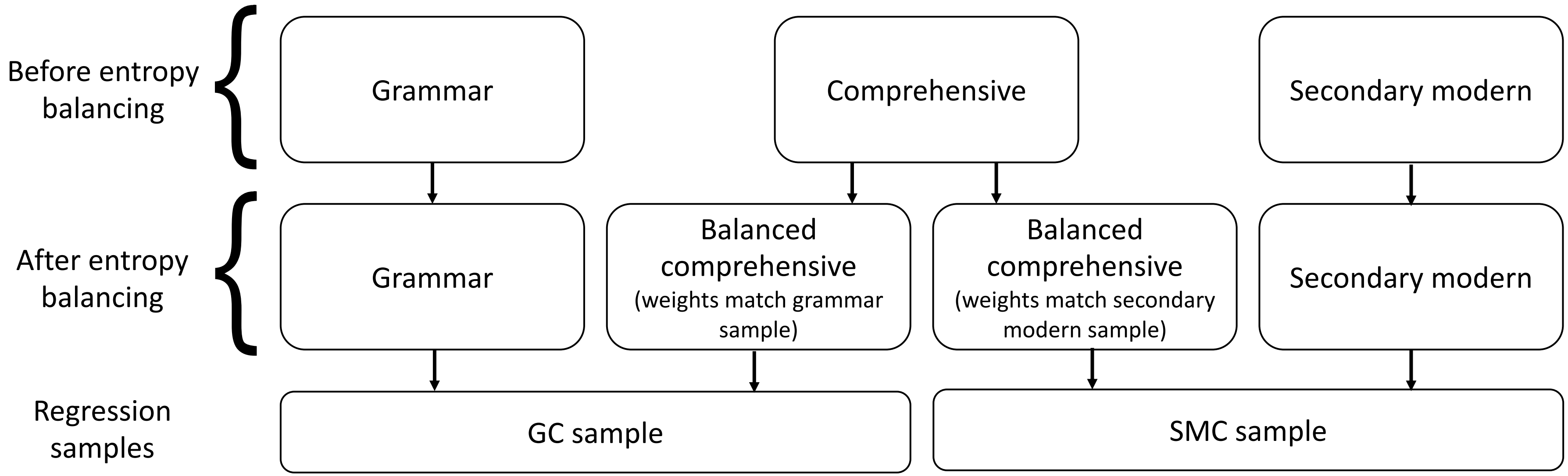


Figure A1: Diagram illustrating working samples before and after the balancing procedure.

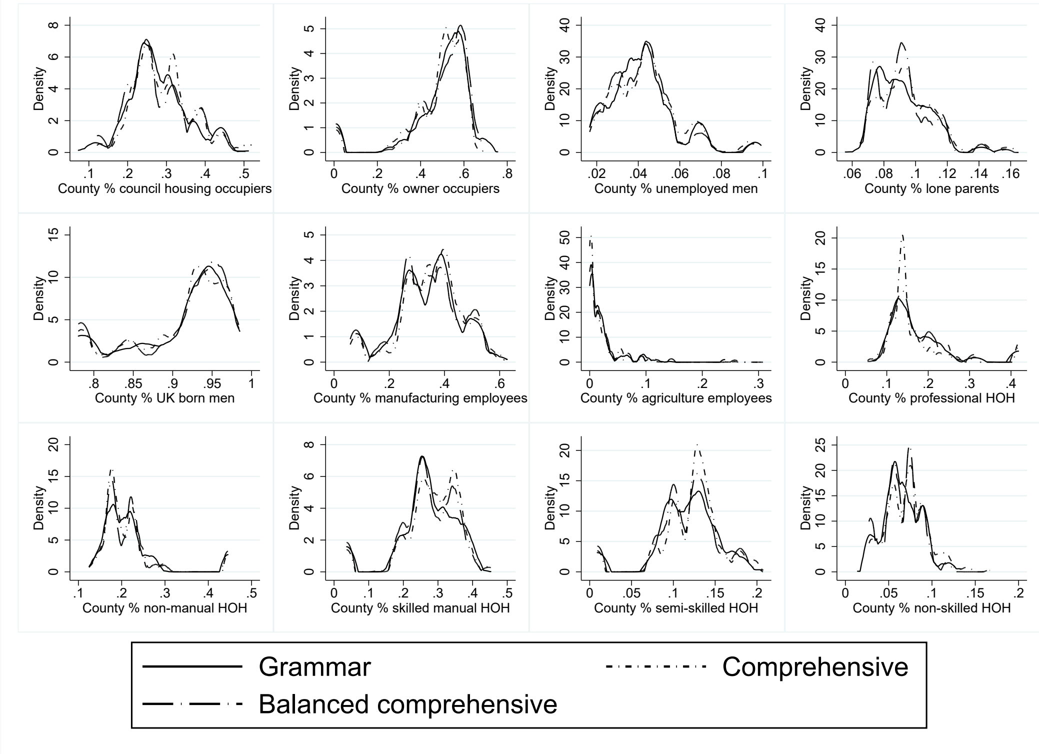


Figure A2: Kernel density estimates for local area characteristics from 1971 Census, for grammar, comprehensive and comprehensive reweighted via entropy balancing weights.

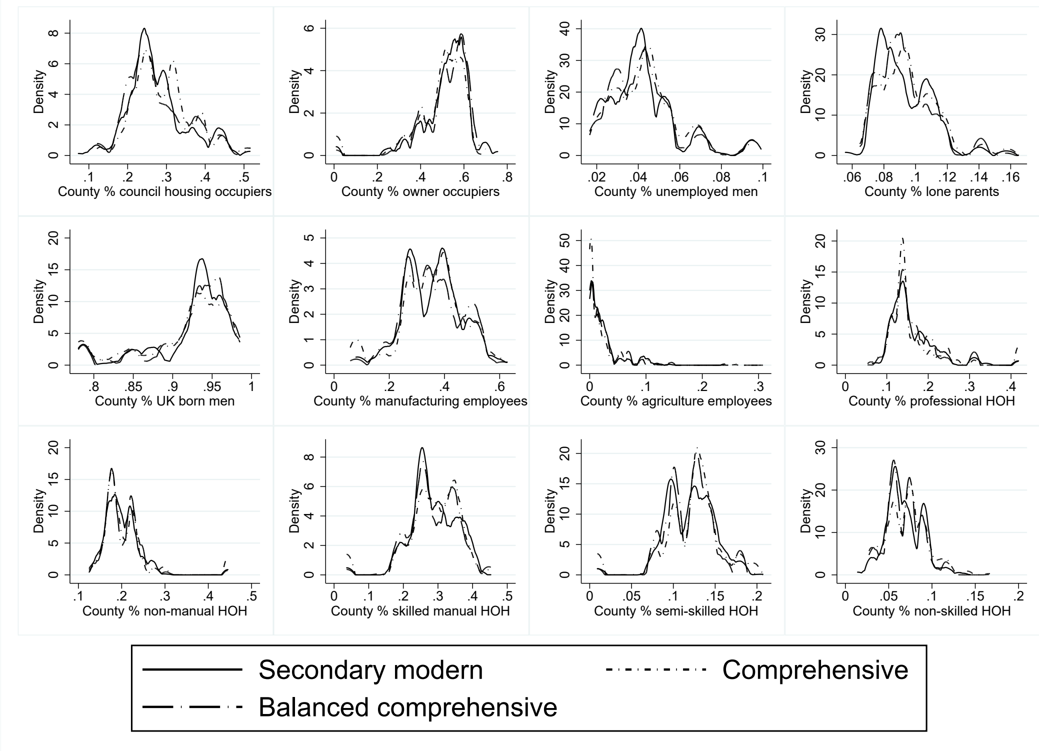


Figure A3: Kernel density estimates for local area characteristics from 1971 Census, for secondary modern, comprehensive and comprehensive reweighed via entropy balancing weights.

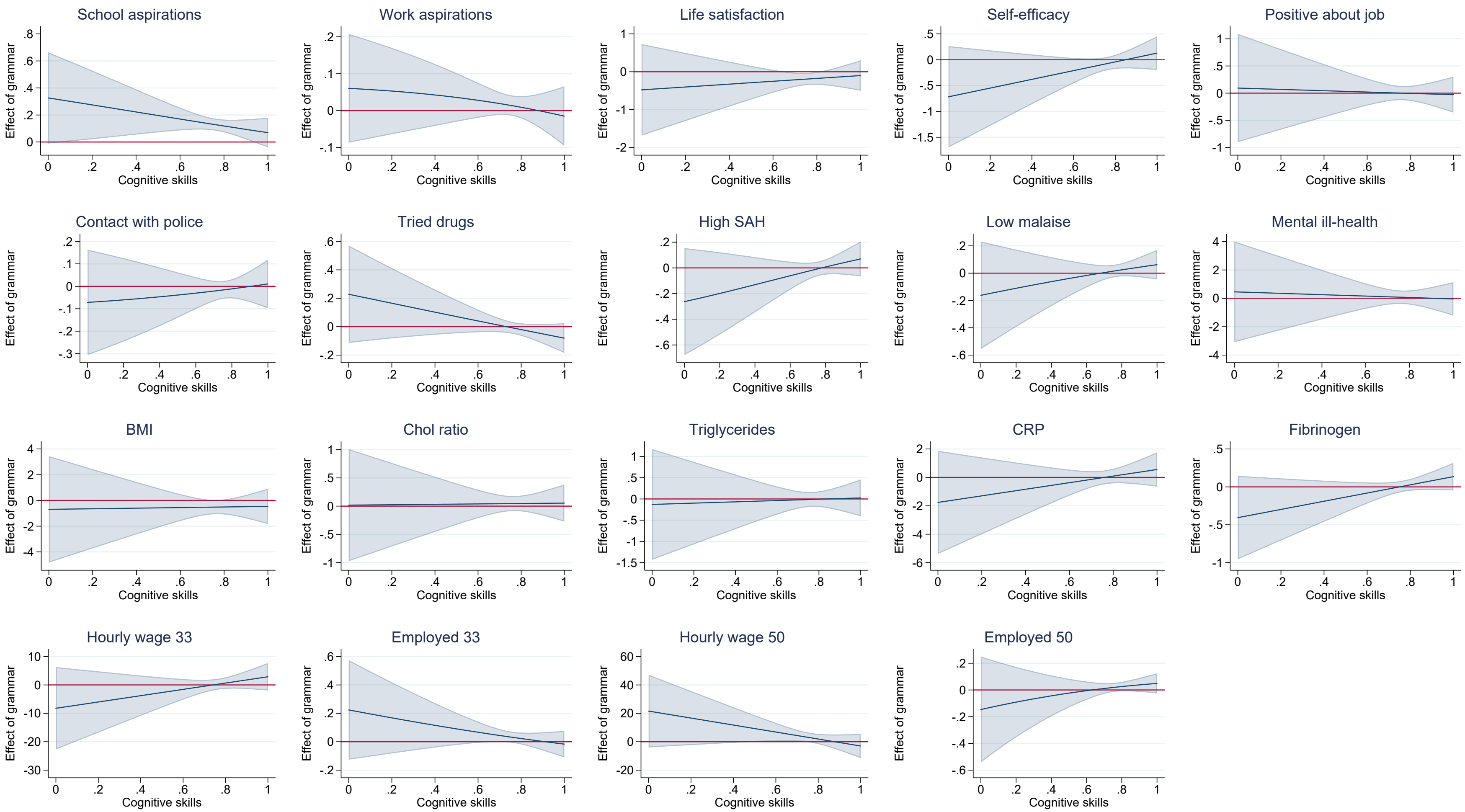


Figure A4: Marginal effect of grammar for different values of age 7 cognitive ability scores. The shaded area represents 95% confidence intervals. Marginal effects are calculated from the interaction between the treatment and ability variable, keeping everything else constant.

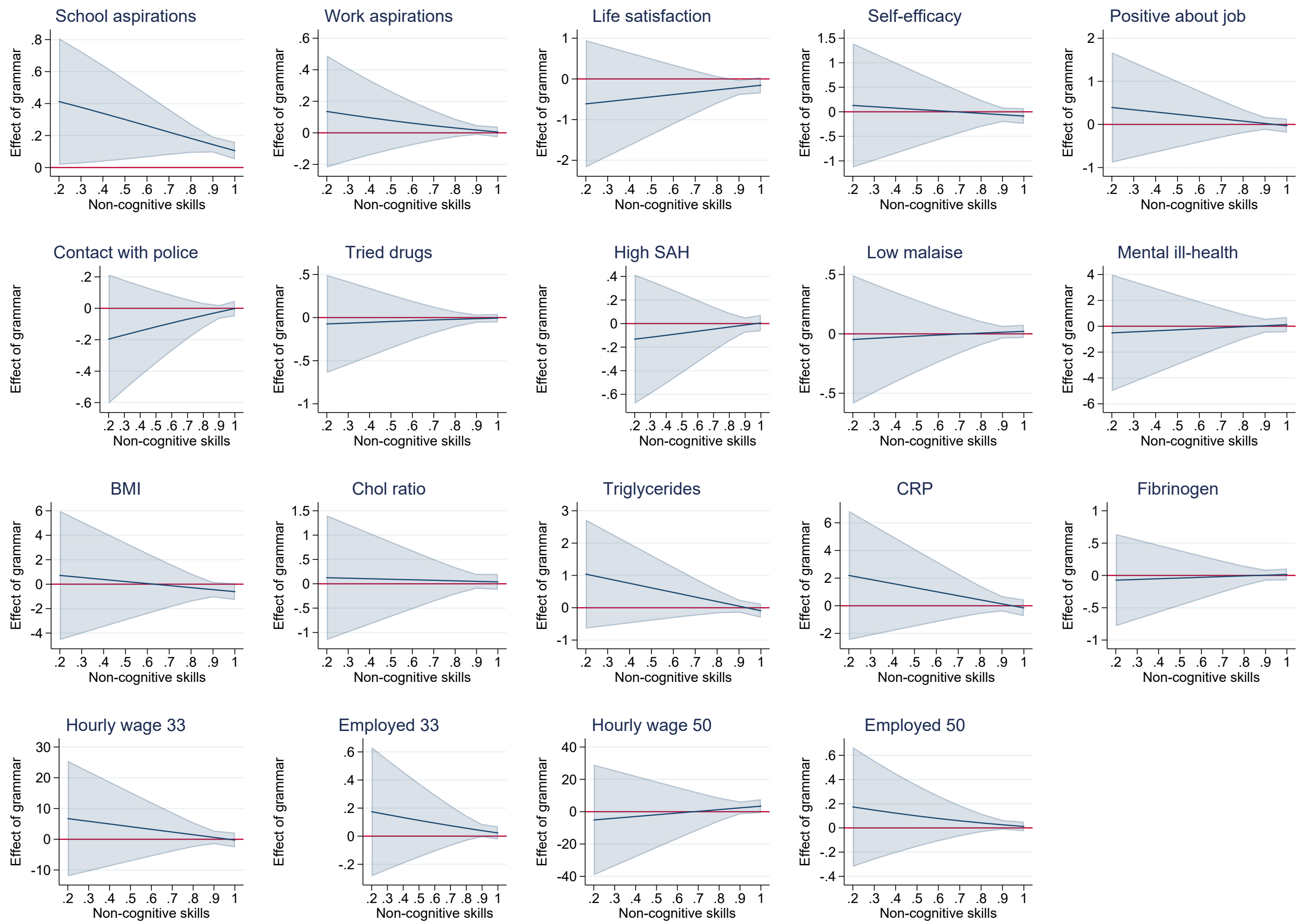


Figure A5: Marginal effect of grammar for different values of non-cognitive ability scores. The shaded area represents 95% confidence intervals. Marginal effects are calculated from the interaction between the treatment and ability variable, keeping everything else constant.

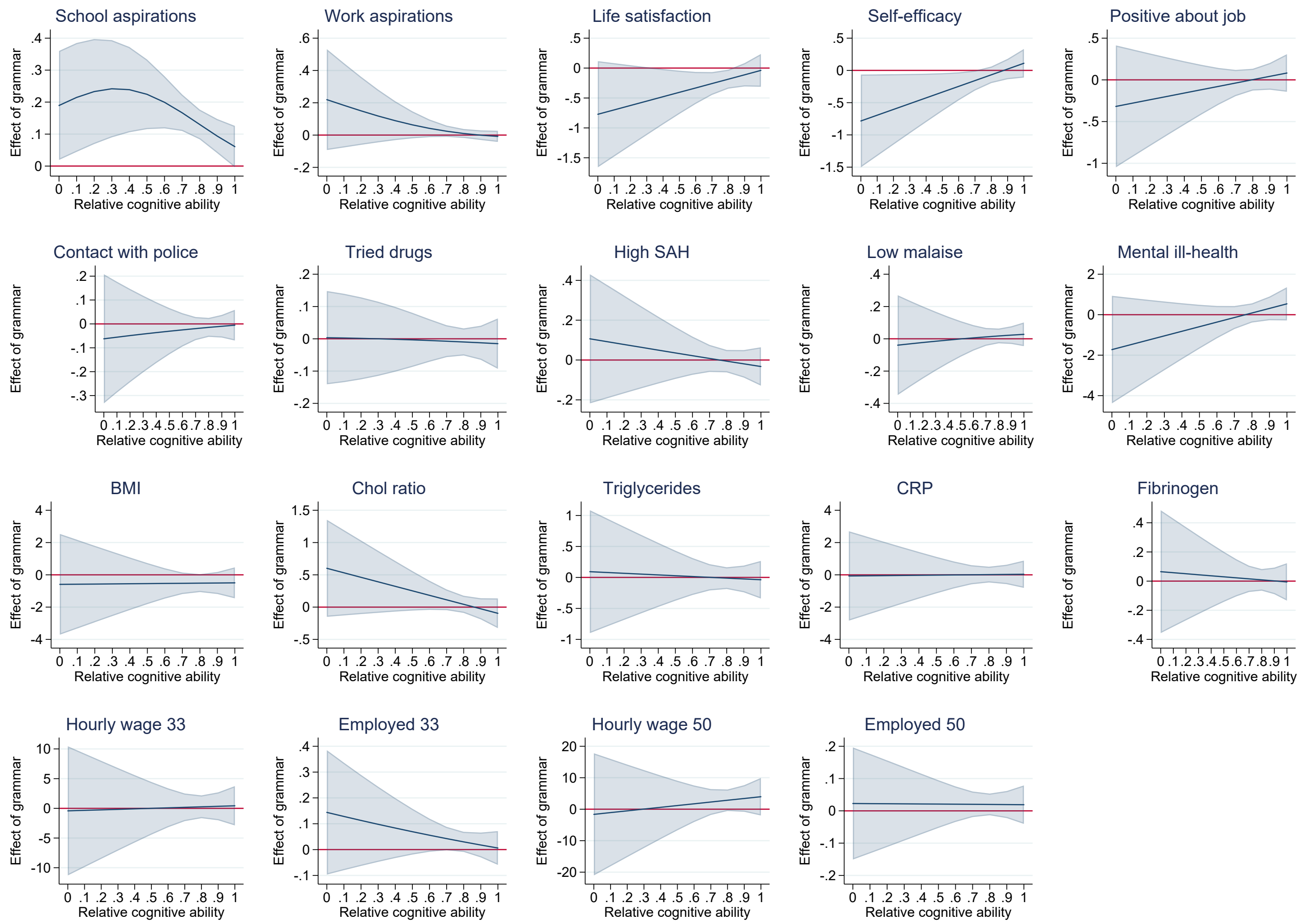


Figure A6: Marginal effect of grammar for different values of age 11 rank of cognitive ability. The shaded area represents 95% confidence intervals. Marginal effects are calculated from the interaction between the treatment and ability variable, keeping everything else constant.

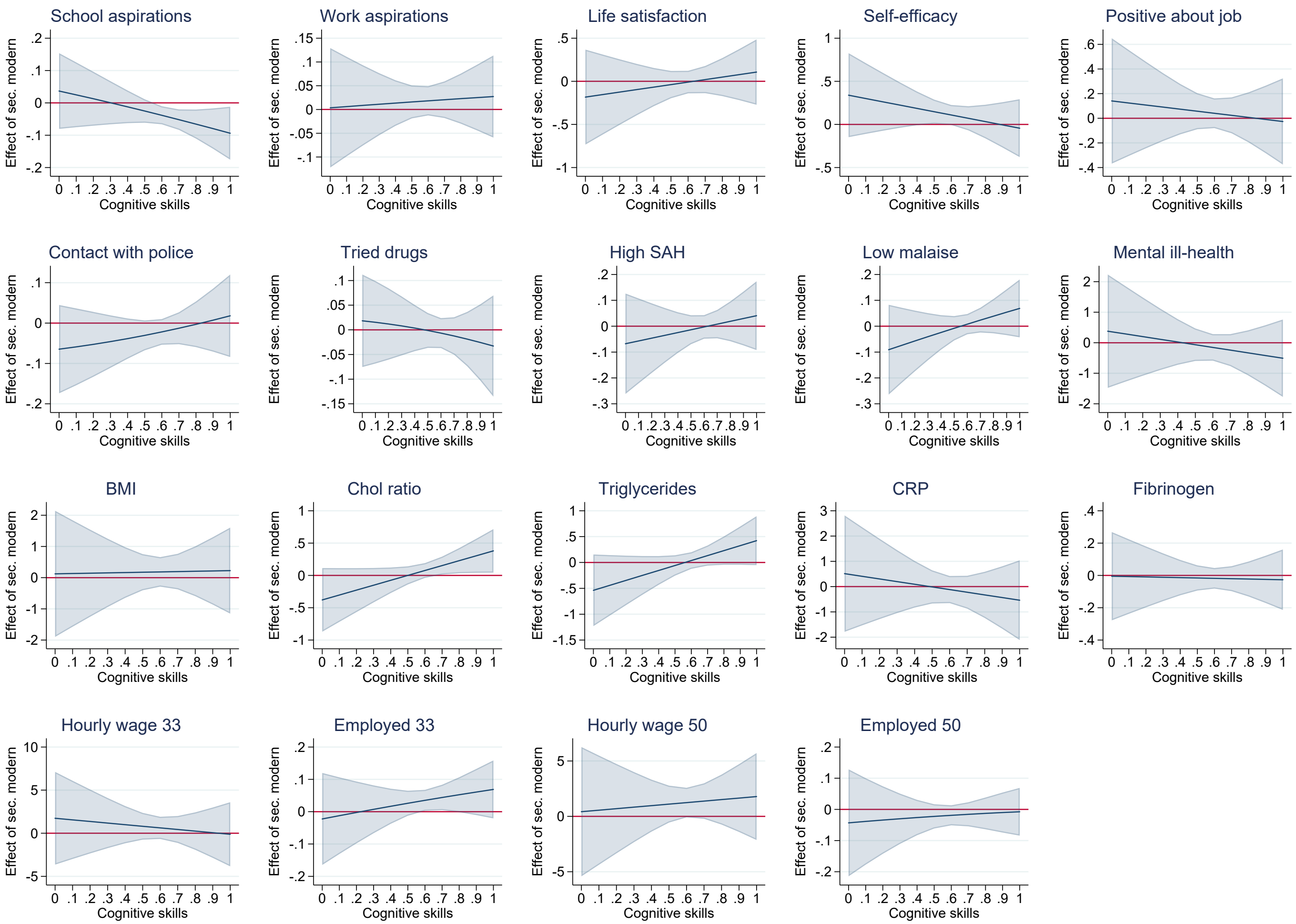


Figure A7: Marginal effect of secondary modern for different values of age 7 cognitive ability scores. The shaded area represents 95% confidence intervals. Marginal effects are calculated from the interaction between the treatment and ability variable, keeping everything else constant.

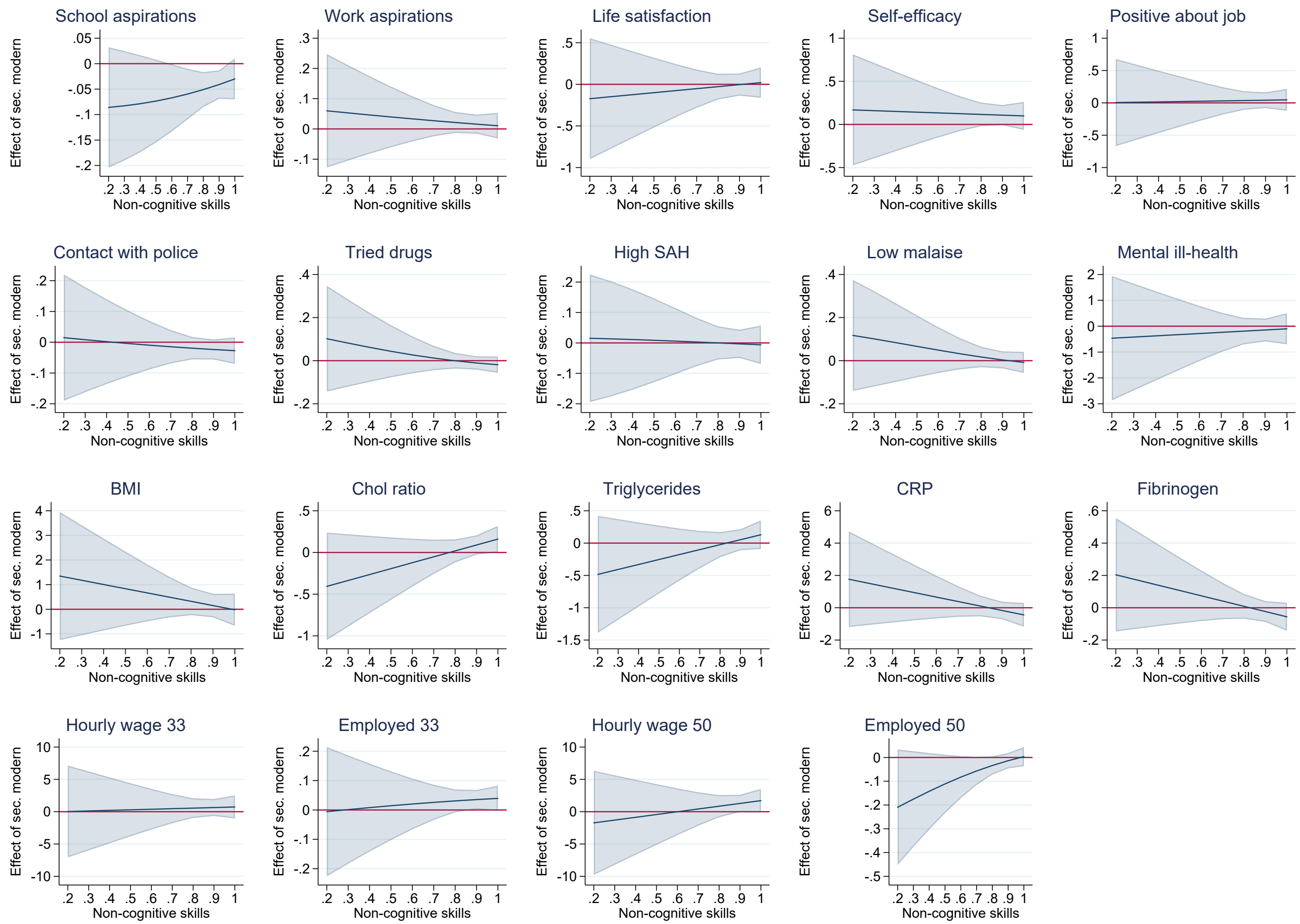


Figure A8: Marginal effect of secondary modern for different values of non-cognitive ability scores. The shaded area represents 95% confidence intervals. Marginal effects are calculated from the interaction between the treatment and ability variable, keeping everything else constant.

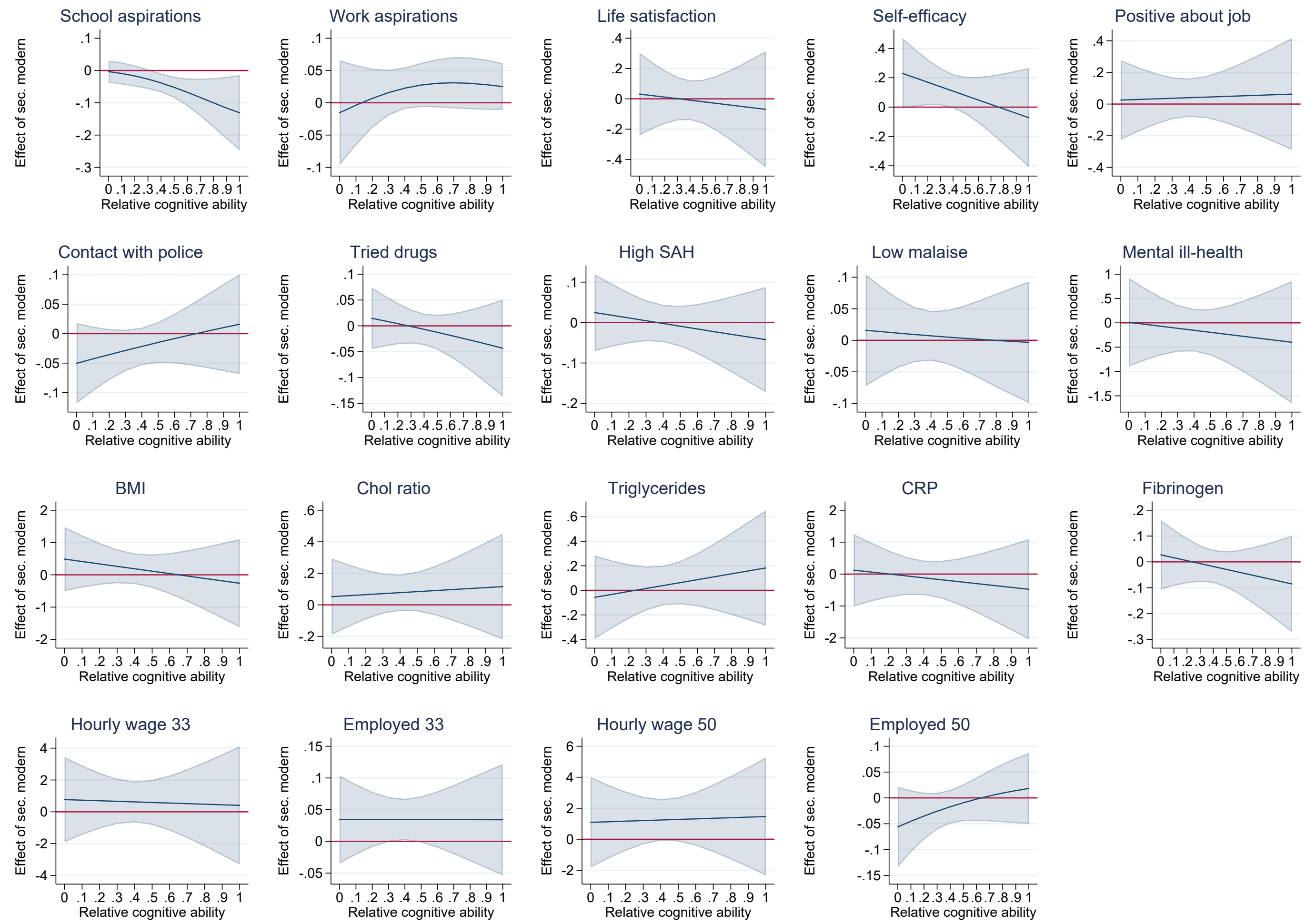


Figure A9: Marginal effect of secondary modern for different values of age 11 rank of cognitive ability. The shaded area represents 95% confidence intervals. Marginal effects are calculated from the interaction between the treatment and ability variable, keeping everything else constant.

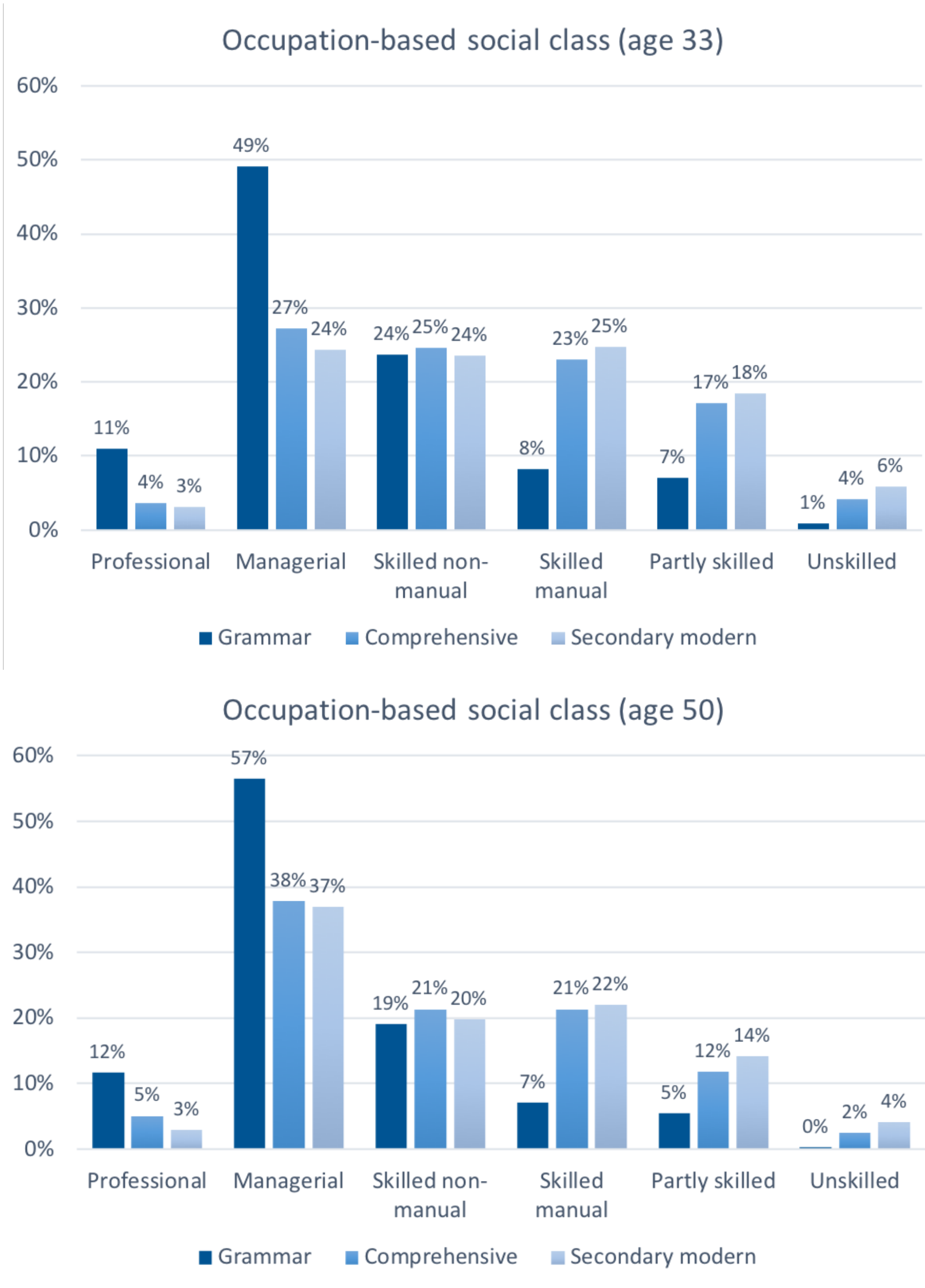


Figure A10: Occupation-based social class by type of secondary school attended. Source: NCDS waves 5 and 8.

A.2 Appendix tables

Table A1: 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 schooling	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

Husband SES is on a 1-5 scale. Mother's schooling is a dummy variable equal to 1 if mother stayed at school after minimum school-leaving age. Maternal smoking during pregnancy is on a scale from 1-Non smoker to 4-Heavy smoker.

Table A2: Durbin-Wu-Hausman test results for the balanced samples.

Grammar and comprehensive sample

	School asp	**Work asp**	**Life sat**	**Self eff**	**Job posit**	**Crime**	**Drugs**	
Wu-Hausman test	0.9647	0.0746	0.0740	0.0322	0.0231	0.2107	3.1718	
p-value	0.3262	0.7847	0.7857	0.8577	0.8792	0.6463	0.0752	
Observations	4159	4156	3131	3083	3145	3277	3279	

	Log wage 33	**Employed 33**	**Log wage 50**	**Employed 50**				
Wu-Hausman test	0.7853	3.7648	0.1339	0.6345				
p-value	0.3757	0.0526	0.7146	0.4259				
Observations	2460	3323	1551	2852				

	SAH	**Low malaise**	**MIH**	**BMI**	**Chol ratio**	**Trig**	**CRP**	**Fib**
Wu-Hausman test	0.0465	1.9587	0.0457	1.1228	0.1072	0.5557	0.1507	0.5885
p-value	0.8294	0.1619	0.8307	0.2895	0.7435	0.4562	0.6979	0.4432
Observations	2875	2854	2787	2759	2327	2333	2302	2295

Secondary modern and comprehensive sample

	School asp	**Work asp**	**Life sat**	**Self eff**	**Job posit**	**Crime**	**Drugs**	
Wu-Hausman test	0.2879	0.2862	0.0000	0.0180	0.2874	0.1548	0.3850	
p-value	0.5916	0.5927	0.9953	0.8933	0.5919	0.6941	0.5350	
Observations	4813	4818	3588	3535	3597	3777	3779	

	Income 33	**Employed 33**	**Income 50**	**Employed 50**				
Wu-Hausman test	1.9379	0.0265	0.5173	1.0000				
p-value	0.1641	0.8708	0.4722	0.3175				
Observations	2766	3821	1689	3230				

	SAH	**Low malaise**	**MIH**	**BMI**	**Chol ratio**	**Trig**	**CRP**	**Fib**
Wu-Hausman test	0.0088	1.5628	0.0046	0.1649	0.4470	0.1729	0.1704	0.1411
p-value	0.9254	0.2114	0.9462	0.6847	0.5039	0.6776	0.6798	0.7072
Observations	3250	3224	3183	3145	2665	2669	2634	2629

The DWH test allows testing for endogeneity in just-identified models. For each outcome, the null hypothesis H_0 is that treatment 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 (i.e. H_0 cannot be rejected), as it is the case for all our outcomes in both samples.

Table A3: Selective schooling and well-being outcomes with treatment interacted with ability.

	School asp.	**Work asp.**	**Life sat.**	**Self-eff.**	**Job posit.**	**Crime**	**Drugs**
Grammar vs comprehensive (high ability)							
Grammar	0.2598***	0.0798*	-0.1094	-0.2095	-0.0028	-0.0535	0.0257
	(0.0498)	(0.0337)	(0.1353)	(0.1618)	(0.1322)	(0.0466)	(0.0430)
Top 50% cognitive skills	0.0432	0.0002	-0.0087	0.0082	-0.0066	-0.0010	0.0412
	(0.0284)	(0.0140)	(0.0680)	(0.0639)	(0.0650)	(0.0248)	(0.0269)
Grammar × top 50% c.s.	-0.1167*	-0.0535	-0.0457	0.1943	0.0374	0.0538	-0.0400
	(0.0469)	(0.0329)	(0.1335)	(0.1580)	(0.1150)	(0.0424)	(0.0452)
Top 50% non-cognitive skills	0.0509+	0.0321*	0.1421*	0.2003**	0.0305	0.0002	-0.0467*
	(0.0270)	(0.0148)	(0.0713)	(0.0664)	(0.0522)	(0.0220)	(0.0224)
Grammar × top 50% n.c.s.	-0.0489	-0.0295	0.0271	-0.0050	-0.0371	-0.0095	-0.0029
	(0.0431)	(0.0275)	(0.1187)	(0.0914)	(0.0915)	(0.0376)	(0.0393)
Observations	4197	4156	3131	3083	3145	3277	3279
F statistic			5.2361	7.2155	15.0213		
χ^2 statistic	889.9119	671.4670				192.4237	157.6848
Secondary modern vs comprehensive (low ability)							
Secondary modern	-0.0449*	0.0113	-0.0477	0.1492*	0.0523	-0.0247	0.0067
	(0.0212)	(0.0199)	(0.0703)	(0.0640)	(0.0588)	(0.0185)	(0.0207)
Top 50% cognitive skills	0.0242	-0.0181	0.0016	0.0754	0.0131	-0.0031	0.0291+
	(0.0198)	(0.0175)	(0.0610)	(0.0499)	(0.0449)	(0.0200)	(0.0172)
Sec. modern × top 50% c.s.	-0.0228	-0.0075	0.0497	-0.0959	-0.0572	0.0217	-0.0161
	(0.0246)	(0.0255)	(0.0812)	(0.0793)	(0.0765)	(0.0254)	(0.0235)
Top 50% non-cognitive skills	0.0050	0.0141	0.1249**	0.1304**	0.0933*	-0.0130	-0.0478***
	(0.0143)	(0.0149)	(0.0455)	(0.0446)	(0.0411)	(0.0168)	(0.0143)
Sec. modern × top 50% n.c.s.	0.0290	0.0200	0.0380	-0.0360	0.0064	-0.0163	-0.0121
	(0.0229)	(0.0250)	(0.0636)	(0.0711)	(0.0737)	(0.0242)	(0.0268)
Observations	4872	4818	3588	3535	3597	3777	3779
F statistic			5.8320	9.5304	21.8894		
χ^2 statistic	1059.3803	669.1582				563.9754	183.4628

+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001. Standard errors clustered at LEA level in parentheses. All control variables are included.

Table A4: Selective schooling and labour market outcomes with treatment interacted with ability.

	Log wage 33	**Employed 33**	**Log wage 50**	**Employed 50**
Grammar vs comprehensive (high ability)				
Grammar	0.2196**	0.0347	0.1264	0.0140
	(0.0819)	(0.0403)	(0.1246)	(0.0403)
Top 50% cognitive skills	0.0707	-0.0020	-0.0473	-0.0017
	(0.0543)	(0.0181)	(0.0616)	(0.0168)
Grammar × top 50% c.s.	-0.1895*	0.0183	0.0178	0.0278
	(0.0944)	(0.0417)	(0.1157)	(0.0344)
Top 50% non-cognitive skills	0.0748	0.0002	0.1836*	0.0125
	(0.0483)	(0.0238)	(0.0928)	(0.0196)
Grammar × top 50% n.c.s.	0.0009	-0.0234	-0.0701	-0.0229
	(0.0776)	(0.0441)	(0.1079)	(0.0312)
Observations	2460	3323	1551	2852
F statistic	28.1158		8.8797	
χ^2 statistic		427.6633		133.1550
Secondary modern vs comprehensive (low ability)				
Secondary modern	-0.0013	0.0352	0.1688**	-0.0344
	(0.0441)	(0.0241)	(0.0588)	(0.0233)
Top 50% cognitive skills	0.0277	-0.0262	-0.0209	0.0244
	(0.0318)	(0.0217)	(0.0592)	(0.0188)
Sec. modern × top 50% c.s.	-0.0028	0.0141	-0.0222	0.0221
	(0.0474)	(0.0296)	(0.0674)	(0.0254)
Top 50% non-cognitive skills	0.0190	0.0221	0.0887	0.0340+
	(0.0281)	(0.0154)	(0.0569)	(0.0187)
Sec. modern × top 50% n.c.s.	0.0971*	-0.0145	-0.1188	0.0128
	(0.0430)	(0.0307)	(0.0760)	(0.0272)
Observations	2766	3821	1689	3230
F statistic	30.3270		19.1561	
χ^2 statistic		394.2505		232.0630

+ p<0.1, * p<0.05, ** p<0.01. Standard errors clustered at LEA level in parentheses. All control variables are included.

Table A5: Selective schooling and health outcomes with treatment interacted with ability.

	SAH	**Low mal.**	**MIH**	**BMI**	**Chol r.**	**Trig**	**CRP**	**Fib**
Grammar vs comprehensive (high ability)								
Grammar	-0.0264	-0.0499	-0.0371	-0.1965	0.0956	0.2788	-0.1559	-0.2421+
	(0.0748)	(0.0543)	(0.1311)	(0.1492)	(0.1664)	(0.2702)	(0.1159)	(0.1431)
Top 50% cognitive skills	-0.0455	-0.0385	-0.0273	-0.0511	0.0180	-0.0027	-0.1242	-0.0941
	(0.0371)	(0.0263)	(0.0584)	(0.0737)	(0.0698)	(0.0688)	(0.1034)	(0.0879)
Grammar × top 50% c.s.	0.0374	0.0671	0.0339	0.0842	-0.0650	-0.1652	0.2883*	0.2834*
	(0.0721)	(0.0565)	(0.1454)	(0.1236)	(0.1301)	(0.1747)	(0.1185)	(0.1202)
Top 50% non-cognitive skills	0.0394	0.0087	-0.0771	-0.0791	-0.0812	0.0094	-0.0082	-0.0946
	(0.0321)	(0.0341)	(0.0595)	(0.0706)	(0.0789)	(0.0546)	(0.0468)	(0.0626)
Grammar × top 50% n.c.s.	-0.0133	0.0141	0.0249	0.0230	-0.0036	-0.1991	-0.1131	0.0236
	(0.0535)	(0.0505)	(0.0957)	(0.1046)	(0.1354)	(0.1620)	(0.0965)	(0.1014)
Observations	2875	2854	2805	2759	2327	2333	2302	2295
F statistic			8.9919	4.9576	59.4320	58.8760	3.1275	4.1532
χ^2 statistic	488.8120	251.0408						
Secondary modern vs comprehensive (low ability)								
Secondary modern	-0.0282	-0.0085	-0.0341	0.0712	-0.0363	-0.0492	0.0109	0.0253
	(0.0320)	(0.0273)	(0.0638)	(0.0846)	(0.0824)	(0.0924)	(0.0544)	(0.0717)
Top 50% cognitive skills	0.0223	-0.0018	0.0259	-0.0076	-0.0529	-0.0650	-0.0187	0.0329
	(0.0281)	(0.0285)	(0.0480)	(0.0663)	(0.0609)	(0.0564)	(0.0758)	(0.0640)
Sec. modern × top 50% c.s.	0.0248	0.0486	-0.0785	0.0037	0.1807*	0.1288	-0.0493	0.0301
	(0.0405)	(0.0383)	(0.0800)	(0.0846)	(0.0804)	(0.0808)	(0.0970)	(0.0844)
Top 50% non-cognitive skills	0.0769**	0.0675**	-0.1514**	0.0559	-0.0577	-0.0465	0.0377	0.0543
	(0.0271)	(0.0209)	(0.0465)	(0.0536)	(0.0512)	(0.0589)	(0.0644)	(0.0489)
Sec. modern × top 50% n.c.s.	0.0300	-0.0086	0.0654	-0.0679	0.0258	0.0184	-0.0206	-0.1238
	(0.0411)	(0.0318)	(0.0675)	(0.0820)	(0.0998)	(0.0865)	(0.1043)	(0.0790)
Observations	3250	3224	3203	3145	2665	2669	2634	2629
F statistic			7.3783	6.3010	15.6410	35.4881	3.4408	5.7191
χ^2 statistic	433.0181	194.4724						

+ p<0.1, * p<0.05, ** p<0.01. Standard errors clustered at LEA level in parentheses. All control variables are included.

Table A6: Selective schooling and well-being outcomes with treatment interacted with sex.

	School asp.	**Work asp.**	**Life sat.**	**Self-eff.**	**Job posit.**	**Crime**	**Drugs**
Grammar vs comprehensive (high ability)							
Grammar	0.1973***	0.0047	-0.0602	0.0733	-0.0409	-0.0327	-0.0082
	(0.0267)	(0.0180)	(0.0814)	(0.0546)	(0.0562)	(0.0241)	(0.0262)
Female	0.0274	0.0035	0.1261+	0.0471	-0.4284***	-0.1266***	-0.0647**
	(0.0229)	(0.0188)	(0.0732)	(0.0614)	(0.0562)	(0.0235)	(0.0210)
Grammar × Female	-0.1279**	0.0157	-0.1286	-0.2337*	0.0762	0.0387	-0.0033
	(0.0429)	(0.0263)	(0.1101)	(0.0908)	(0.0912)	(0.0378)	(0.0408)
Observations	4197	4156	3131	3083	3145	3277	3279
F statistic			5.3445	8.0817	15.4014		
χ^2 statistic	883.5693	535.0712				180.7076	140.5953
Secondary modern vs comprehensive (low ability)							
Secondary modern	-0.0069	0.0321	-0.0622	0.0717	0.0483	-0.0072	0.0015
	(0.0185)	(0.0205)	(0.0538)	(0.0563)	(0.0445)	(0.0167)	(0.0193)
Female	0.0531***	0.0095	0.0986+	-0.0335	-0.5588***	-0.1765***	-0.0741***
	(0.0145)	(0.0181)	(0.0550)	(0.0583)	(0.0505)	(0.0170)	(0.0143)
Sec. modern × Female	-0.0670**	-0.0295	0.1089	0.0199	-0.0411	-0.0412	-0.0181
	(0.0223)	(0.0288)	(0.0834)	(0.0933)	(0.0841)	(0.0266)	(0.0245)
Observations	4872	4818	3588	3535	3597	3777	3779
F statistic			5.5894	10.7201	19.7839		
χ^2 statistic	1047.1880	698.3471				580.8757	174.4167

+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001. Standard errors clustered at LEA level in parentheses. All continuous outcomes are standardised. Binary outcomes are estimated via probit models, for which marginal effects are displayed. All control variables are included.

Table A7: Selective schooling and labour market outcomes with treatment interacted with sex.

	Log wage 33	**Employed 33**	**Log wage 50**	**Employed 50**
Grammar vs comprehensive (high ability)				
Grammar	0.1133*	0.1076*	0.1462*	0.0245
	(0.0451)	(0.0444)	(0.0600)	(0.0243)
Female	-0.3759***	-0.2316***	-0.1663**	-0.0621***
	(0.0518)	(0.0240)	(0.0615)	(0.0184)
Grammar × Female	-0.1059	-0.0924+	-0.1088	-0.0071
	(0.0774)	(0.0545)	(0.0750)	(0.0282)
Observations	2460	3323	1551	2852
F statistic	26.8309		9.6014	
χ^2 statistic		360.4620		125.1572
Secondary modern vs comprehensive (low ability)				
Secondary modern	0.0704+	0.0706*	0.1563**	-0.0043
	(0.0379)	(0.0274)	(0.0584)	(0.0216)
Female	-0.4565***	-0.2228***	-0.2018***	-0.0682***
	(0.0297)	(0.0216)	(0.0444)	(0.0173)
Sec. modern × Female	-0.0438	-0.0538+	-0.1329+	-0.0261
	(0.0593)	(0.0318)	(0.0712)	(0.0315)
Observations	2766	3821	1689	3230
F statistic	29.8599		21.7543	
χ^2 statistic		410.8919		237.0804

+ $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors clustered at LEA level in parentheses. Binary outcomes are estimated via probit models, for which marginal effects are displayed. All control variables are included.

Table A8: Selective schooling and health outcomes with treatment interacted with sex.

	SAH	**Low mal.**	**MIH**	**BMI**	**Chol r.**	**Trig**	**CRP**	**Fib**
Grammar vs comprehensive (high ability)								
Grammar	-0.0355	0.0909**	-0.0255	-0.0719	0.1472	0.0089	0.1404*	0.0771
	(0.0399)	(0.0341)	(0.0622)	(0.0648)	(0.0900)	(0.0874)	(0.0640)	(0.0664)
Female	-0.0123	-0.0387+	0.1936***	-0.2020**	-0.7591***	-0.6610***	0.1346*	0.2789***
	(0.0319)	(0.0225)	(0.0528)	(0.0726)	(0.0719)	(0.0694)	(0.0515)	(0.0711)
Grammar × Female	0.0561	-0.1208**	0.0677	-0.0632	-0.1999+	-0.0292	-0.2544**	-0.1163
	(0.0513)	(0.0391)	(0.0892)	(0.1025)	(0.1031)	(0.0924)	(0.0770)	(0.1051)
Observations	2875	2854	2805	2759	2327	2333	2302	2295
F statistic			8.9614	5.0908	64.9295	52.2225	4.2681	3.7355
χ^2 statistic	496.3597	250.2092						
Secondary modern vs comprehensive (low ability)								
Secondary modern	-0.0001	-0.0057	0.0500	-0.0055	0.0856	0.0351	-0.0074	0.0640
	(0.0328)	(0.0247)	(0.0518)	(0.0576)	(0.0607)	(0.0759)	(0.0941)	(0.0587)
Female	-0.0346	-0.1268***	0.3495***	-0.2071***	-0.6450***	-0.5310***	0.1360*	0.3605***
	(0.0247)	(0.0223)	(0.0468)	(0.0468)	(0.0570)	(0.0577)	(0.0634)	(0.0497)
Sec. modern × Female	-0.0032	0.0229	-0.1639*	0.0791	-0.0333	-0.0189	-0.0357	-0.1774*
	(0.0419)	(0.0328)	(0.0671)	(0.0760)	(0.0811)	(0.0864)	(0.1200)	(0.0777)
Observations	3250	3224	3203	3145	2665	2669	2634	2629
F statistic			7.6761	6.0386	16.2085	25.7998	3.4973	6.4892
χ^2 statistic	434.5547	207.6119						

+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001. Standard errors clustered at LEA level in parentheses. All continuous outcomes are standardised. Binary outcomes are estimated via probit models, for which marginal effects are displayed. All control variables are included.

Table A9: Selective schooling and well-being outcomes, distinguishing between comprehensives by origin.

	School asp.	**Work asp.**	**Life sat.**	**Self-eff.**	**Job positiv.**	**Crime**	**Drugs**
Grammar vs comprehensive (high ability)							
Comprehensive (former grammar)	-0.1023***	-0.0209	0.1479*	0.0506	0.0475	0.0354	0.0033
	(0.0279)	(0.0190)	(0.0684)	(0.0524)	(0.0616)	(0.0245)	(0.0334)
Comprehensive (other)	-0.1336***	-0.0103	0.1241*	0.0547	-0.0163	0.0060	0.0129
	(0.0188)	(0.0114)	(0.0513)	(0.0416)	(0.0485)	(0.0190)	(0.0185)
Observations	4197	4156	3131	3083	3145	3277	3279
F statistic			5.4937	7.6043	15.4003		
χ^2 statistic	305.40	52.50				64.39	66.08
Secondary modern vs comprehensive (low ability)							
Comprehensive (former sec modern)	0.0201	-0.0385+	0.0461	-0.0547	-0.0069	0.0052	-0.0194
	(0.0157)	(0.0216)	(0.0657)	(0.0537)	(0.0573)	(0.0180)	(0.0187)
Comprehensive (other)	0.0495***	-0.0093	-0.0082	-0.0883*	-0.0331	0.0268+	0.0162
	(0.0141)	(0.0155)	(0.0450)	(0.0380)	(0.0398)	(0.0146)	(0.0157)
Observations	4872	4818	3588	3535	3597	3777	3779
F statistic			5.6525	10.1448	20.0561		
χ^2 statistic	354.47	276.89				207.54	86.04

+ p<0.1, * p<0.05, *** p<0.001. Standard errors clustered at LEA level in parentheses. All continuous outcomes are standardised. Binary outcomes are estimated via probit models, for which marginal effects are displayed. All control variables are included.

Table A10: Selective schooling and labour market outcomes, distinguishing between comprehensives by origin.

	Log hourly wage 33	**Employed at 33**	**Log hourly wage 50**	**Employed at 50**
Grammar vs comprehensive (high ability)				
Comprehensive (former grammar)	-0.0616	0.0048	-0.0678	0.0144
	(0.0470)	(0.0262)	(0.1011)	(0.0209)
Comprehensive (other)	-0.0578+	-0.0463**	-0.0962*	-0.0308*
	(0.0342)	(0.0164)	(0.0450)	(0.0150)
Observations	2460	3323	1551	2852
F statistic	26.6779		9.5117	
χ^2 statistic		191.98		50.09
Secondary modern vs comprehensive (low ability)				
Comprehensive (former sec modern)	-0.0209	-0.0334	-0.0911+	0.0267+
	(0.0434)	(0.0204)	(0.0536)	(0.0161)
Comprehensive (other)	-0.0581+	-0.0340*	-0.0791+	0.0181
	(0.0321)	(0.0157)	(0.0417)	(0.0157)
Observations	2766	3821	1689	3230
F statistic	29.1982		20.3900	
χ^2 statistic		267.75		109.31

+ p<0.1, * p<0.05, ** p<0.01. Standard errors clustered at LEA level in parentheses. Binary outcomes are estimated via probit models, for which marginal effects are displayed. All control variables are included.

Table A11: Selective schooling and health outcomes, distinguishing between comprehensives by origin.

	High SAH	**Low malaise**	**Mental ill-health**	**BMI**	**Chol ratio**	**Triglycerides**	**CRP**	**Fibrinogen**
Grammar vs comprehensive (high ability)								
Comprehensive (former grammar)	-0.0023	0.0104	0.0014	0.0419	-0.1452*	-0.0413	-0.0415	-0.1660*
	(0.0369)	(0.0339)	(0.0685)	(0.0703)	(0.0590)	(0.0603)	(0.0516)	(0.0832)
Comprehensive (other)	0.0085	-0.0272	-0.0176	0.1310*	0.0045	0.0263	0.0130	0.0507
	(0.0333)	(0.0206)	(0.0523)	(0.0551)	(0.0590)	(0.0601)	(0.0489)	(0.0589)
Observations	2875	2854	2805	2759	2327	2333	2302	2295
F statistic			9.4048	4.9457	47.2419	40.2559	3.4493	4.7212
χ^2 statistic	52.9100	57.4897						
Secondary modern vs comprehensive (low ability)								
Comprehensive (former sec modern)	0.0111	-0.0282	-0.0333	-0.1043+	-0.1045	-0.0012	-0.0836+	-0.0893
	(0.0320)	(0.0292)	(0.0557)	(0.0622)	(0.0634)	(0.0626)	(0.0475)	(0.0659)
Comprehensive (other)	0.0003	0.0031	0.0554	-0.0077	-0.0525	-0.0334	0.0657	0.0747+
	(0.0234)	(0.0185)	(0.0395)	(0.0491)	(0.0459)	(0.0476)	(0.0527)	(0.0441)
Observations	3250	3224	3203	3145	2665	2669	2634	2629
F statistic			5.9683	6.1660	16.5526	27.5086	5.8011	6.2885
χ^2 statistic	109.3500	97.7415						

+ p<0.1, * p<0.05. Standard errors clustered at LEA level in parentheses. All continuous outcomes are standardised. Binary outcomes are estimated via probit models, for which marginal effects are displayed. All control variables are included.

Table A12: Selective schooling and well-being outcomes, using only completely selective or completely comprehensive LEAs.

	School asp.	**Work asp.**	**Life sat.**	**Self-eff.**	**Job positiv.**	**Crime**	**Drugs**
Grammar vs comprehensive (high ability)							
Grammar	0.2213***	0.0156	-0.0821	0.0449	-0.1703	0.0018	0.0337
	(0.0434)	(0.0370)	(0.1205)	(0.1192)	(0.1299)	(0.0701)	(0.0498)
Observations	860	842	640	635	642	678	678
Secondary modern vs comprehensive (low ability)							
Secondary modern	-0.0115	0.0547+	0.0867	0.2197**	-0.0581	0.0378+	-0.0099
	(0.0254)	(0.0317)	(0.0904)	(0.0739)	(0.0740)	(0.0224)	(0.0371)
Observations	1006	993	756	747	753	790	786

+ p<0.1, ** p<0.01, *** p<0.001. Standard errors clustered at LEA level in parentheses. All continuous outcomes are standardised. Binary outcomes are estimated via probit models, for which marginal effects are displayed. All control variables are included.

Table A13: Selective schooling and labour market outcomes, using only completely selective or completely comprehensive LEAs.

	Log hourly wage 33	**Employed at 33**	**Log hourly wage 50**	**Employed at 50**
Grammar vs comprehensive (high ability)				
Grammar	0.0475	0.0914+	0.3275+	0.0209
	(0.0680)	(0.0549)	(0.1875)	(0.0476)
Observations	508	680	294	562
Secondary modern vs comprehensive (low ability)				
Secondary modern	0.0358	0.0503+	0.0636	0.0135
	(0.0764)	(0.0281)	(0.0915)	(0.0290)
Observations	592	796	330	642

+ p<0.1. Standard errors clustered at LEA level in parentheses. Binary outcomes are estimated via probit models, for which marginal effects are displayed. All control variables are included.

Table A14: Selective schooling and health outcomes, using only completely selective or completely comprehensive LEAs.

	High SAH	**Low malaise**	**Mental ill-health**	**BMI**	**Chol ratio**	**Triglycerides**	**CRP**	**Fibrinogen**
Grammar vs comprehensive (high ability)								
Grammar	0.0250	0.0281	-0.0477	-0.4470***	-0.0267	-0.0379	-0.3291***	-0.3463**
	(0.0592)	(0.0692)	(0.1302)	(0.1268)	(0.1416)	(0.1235)	(0.0748)	(0.1144)
Observations	581	575	575	561	449	449	441	442
Secondary modern vs comprehensive (low ability)								
Secondary modern	0.0113	0.0242	-0.0627	0.1167	0.0056	0.0583	-0.0402	0.0125
	(0.0497)	(0.0428)	(0.1136)	(0.0946)	(0.0919)	(0.0903)	(0.1047)	(0.1663)
Observations	658	650	648	635	515	515	506	506

** p<0.01, *** p<0.001. Standard errors clustered at LEA level in parentheses. All continuous outcomes are standardised. Binary outcomes are estimated via probit models, for which marginal effects are displayed. All control variables are included.

Table A15: First stage for grammar and secondary modern attendance.

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

*** $p<0.001$. Standard errors clustered at LEA level in parentheses. The IV is % pupils going to comprehensive schools in individual's LEA.

Table A16: 2SLS estimates for well-being outcomes.

	School	**Work**	**Life sat.**	**Self-eff.**	**Job positiv.**	**Crime**	**Drugs**
Grammar vs comprehensive (high ability)							
Grammar	0.1115**	0.0028	-0.1621	-0.0348	-0.0137	-0.0324	0.0725+
	(0.0421)	(0.0292)	(0.1106)	(0.1045)	(0.1145)	(0.0347)	(0.0408)
Cognitive skills	0.1262	-0.0783	-0.0527	0.5896**	0.2403	0.0750	-0.0178
	(0.0935)	(0.0610)	(0.2753)	(0.2141)	(0.2446)	(0.0958)	(0.0921)
Non-cognitive skills	0.1749	0.0753	1.1859**	1.1129***	0.0421	-0.0502	-0.3342**
	(0.1224)	(0.0728)	(0.3625)	(0.3226)	(0.2598)	(0.1099)	(0.1188)
Relative cogn. ability	0.7193***	0.2032***	-0.1181	0.3586*	0.8331***	-0.0791	0.1563**
	(0.0675)	(0.0518)	(0.1588)	(0.1615)	(0.1548)	(0.0619)	(0.0590)
Observations	4197	4156	3131	3083	3145	3277	3279
F statistic	24.3526	5.9191	4.6924	7.5238	13.4001	5.0295	4.6612
Secondary modern vs comprehensive (low ability)							
Secondary modern	-0.0571*	0.0536*	-0.0006	0.0635	-0.0423	-0.0251	0.0071
	(0.0256)	(0.0258)	(0.0812)	(0.0704)	(0.0707)	(0.0282)	(0.0272)
Cognitive skills	-0.0000	0.0651	0.2016	0.3829+	0.2521	0.0887	0.1085*
	(0.0499)	(0.0617)	(0.1583)	(0.1957)	(0.1558)	(0.0568)	(0.0530)
Non-cognitive skills	0.0928+	0.1632*	0.5836***	0.3908*	0.4616**	-0.1403*	-0.2607***
	(0.0472)	(0.0645)	(0.1631)	(0.1600)	(0.1434)	(0.0637)	(0.0681)
Relative cogn. ability	0.3847***	0.3920***	-0.2026*	0.2251*	0.6089***	-0.0542	0.0347
	(0.0427)	(0.0427)	(0.0958)	(0.1100)	(0.1168)	(0.0432)	(0.0414)
Observations	4872	4818	3588	3535	3597	3777	3779
F statistic	36.1403	16.9909	5.7924	9.6062	20.1324	18.3461	4.5302

+ p<0.1, * p<0.05. The IV is percentage of comprehensive pupils in the individual's LEA. Standard errors clustered at LEA level in parentheses. All continuous outcomes are standardised. All control variables are included.

Table A17: 2SLS estimates for labour market outcomes.

	Log hourly wage 33	**Employed at 33**	**Log hourly wage 50**	**Employed at 50**
Grammar vs comprehensive (high ability)				
Grammar	0.1565*	0.1063**	0.0556	0.0453
	(0.0688)	(0.0349)	(0.0969)	(0.0323)
Cognitive skills	0.0001	0.0837	-0.0064	0.0670
	(0.1662)	(0.0990)	(0.2769)	(0.0911)
Non-cognitive skills	0.4243*	0.0291	0.7548*	0.1031
	(0.2033)	(0.1076)	(0.2991)	(0.0877)
Relative cogn. ability	0.6627***	0.0783	0.7927***	-0.0074
	(0.1203)	(0.0707)	(0.1542)	(0.0508)
Observations	2460	3323	1551	2852
F statistic	25.2729	16.2169	8.0179	3.7896
Secondary modern vs comprehensive (low ability)				
Secondary modern	-0.0546	0.0261	-0.0179	-0.0562+
	(0.0583)	(0.0267)	(0.0771)	(0.0318)
Cognitive skills	0.1859*	0.0483	0.0595	0.1214*
	(0.0832)	(0.0640)	(0.1325)	(0.0570)
Non-cognitive skills	0.2705*	0.1230*	-0.0160	0.2118**
	(0.1117)	(0.0514)	(0.1858)	(0.0658)
Relative cogn. ability	0.3624***	0.0502	0.4460***	0.1076**
	(0.0631)	(0.0403)	(0.0825)	(0.0359)
Observations	2766	3821	1689	3230
F statistic	31.6327	13.4527	18.0886	6.6611

+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001. The IV is percentage of comprehensive pupils in the individual's LEA. Standard errors clustered at LEA level in parentheses. All continuous outcomes are standardised. All control variables are included.

Table A18: 2SLS estimates for health outcomes.

	SAH	**Low mal.**	**MIH**	**BMI**	**Chol r.**	**Trig**	**CRP**	**Fib**
Grammar vs comprehensive (high ability)								
Grammar	0.0076	0.0802+	0.0256	-0.2357*	-0.0021	-0.0957	-0.0834	-0.0376
	(0.0631)	(0.0450)	(0.0987)	(0.1059)	(0.1202)	(0.0917)	(0.0940)	(0.1079)
Cognitive skills	0.2238+	0.0723	-0.1827	-0.6264**	-0.5397+	-0.5752+	-0.1335	-0.0611
	(0.1270)	(0.1141)	(0.2923)	(0.2336)	(0.3030)	(0.2973)	(0.1975)	(0.2835)
Non-cognitive skills	0.2247	0.1004	-0.6175*	-0.4060	-0.3511	-0.2466	-0.2144	-0.5380+
	(0.1613)	(0.1245)	(0.2927)	(0.2829)	(0.3538)	(0.3497)	(0.2157)	(0.3126)
Relative cogn. ability	0.0970	0.1216	0.2424	-0.0065	-0.2198	-0.0850	-0.4713*	-0.4247+
	(0.1163)	(0.0810)	(0.1820)	(0.2020)	(0.1819)	(0.2484)	(0.1920)	(0.2399)
Observations	2875	2854	2805	2759	2327	2333	2302	2295
F statistic	16.1225	9.0062	9.6329	4.6837	56.2504	49.0434	3.6831	3.7375
Observations	2875	2854	2787	2759	2327	2333	2302	2295
F statistic	1.3807	1.5620	1.4590	2.0726	7.1932	5.6424	0.5509	1.5584
Secondary modern vs comprehensive (low ability)								
Secondary modern	-0.0040	-0.0278	-0.0568	-0.0220	0.1321	0.0888	-0.0015	-0.0192
	(0.0394)	(0.0345)	(0.0866)	(0.0981)	(0.0900)	(0.0805)	(0.1019)	(0.0937)
Cognitive skills	0.0610	0.0376	-0.1484	-0.0853	-0.0251	-0.0820	-0.2522	-0.0478
	(0.0834)	(0.0771)	(0.1622)	(0.1882)	(0.1989)	(0.2210)	(0.2130)	(0.1734)
Non-cognitive skills	0.3458***	0.3672***	-0.6105***	0.1165	-0.2670	-0.2153	0.1944	0.0906
	(0.0853)	(0.0922)	(0.1635)	(0.1828)	(0.2482)	(0.2217)	(0.2027)	(0.1959)
Relative cogn. ability	0.1534**	0.1248**	-0.1606	-0.1419	-0.0984	-0.0799	-0.1593	-0.4418***
	(0.0549)	(0.0463)	(0.1220)	(0.1327)	(0.1296)	(0.1090)	(0.1466)	(0.1178)
Observations	3250	3224	3203	3145	2665	2669	2634	2629
F statistic	11.9297	6.1111	5.8470	6.0740	16.5835	27.6837	4.0834	5.7843

+ p<0.1, * p<0.05, ** p<0.01, *** p<0.001. The IV is percentage of comprehensive pupils in the individual's LEA. Standard errors clustered at LEA level in parentheses. All continuous outcomes are standardised. All control variables are included.

A.3 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 shown in the first table below. As a general rule (Kaiser's rule), components are retained if their associated eigenvalue exceeds 1. For the three indices, this was only the case for the first component. The eigenvalues obtained by 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.

	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 16	1.0000			10,536
Reading 16	0.6552	1.0000		10,596

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.

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.

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 the age 16 tests have the same factor loading, and the first component, with eigenvalue 1.66, explains 0.83 of the variance.

A.4 Details on outcome variables

Outcomes constructed by PCA

- Age 33 life satisfaction. Cohort members were asked to rate the following: Happiness, all things considered (Not at all to Very), Satisfaction with way life has turned out so far (0-10), Expected satisfaction with life in 10 years (0-10).

- Age 33 self-efficacy. Cohort members were asked to say True or False to the following: I never get what I want of life, I usually have control over life, I can run my life as I want.

- Age 33 positive feelings about one's job. Cohort members were asked to rate the following from 1 to 5: I hang on to a job even if I don't really like it, Work requires me to keep learning new things, Work is monotonous because always do same things, Present work skills will be valuable in 5 years.

9-item Malaise Inventory

- Do you feel tired most of the time?
- Do you often feel miserable or depressed?
- Do you often get worried about things?
- Do you often get into a violent rage?
- Do you often suddenly become scared for no reason?
- Are you easily upset or irritated?
- Are you constantly keyed up and jittery?
- Does every little thing get on your nerves?
- Does your heart often race like mad?

Sources: Ploubidis et al. (2017) and Rutter et al. (1970).

A.5 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:

$$Y_{16i} = \beta_0 + \beta_1 A_{16i} + \beta_2 Comp_i + \beta_3 B_i + \epsilon_{16i}. \tag{A.1}$$

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):

$$Y_{16i} = \alpha_0 + \alpha_1 Y_{11i} + \alpha_2 Comp_i + \alpha_3 B_i + \eta_{16i}. \tag{A.2}$$

A similar specification is assumed to hold for pre-secondary school educational outcomes at age 11. As a 'falsification test', Manning and Pischke (2006) control for comprehensive attendance, which should presumably not be a predictor of pre-secondary school outcomes:

$$Y_{11i} = \gamma_0 + \gamma_1 Y_{7i} + \gamma_2 Comp_i + \gamma_3 B_i + \epsilon_{11i}. \tag{A.3}$$

If the estimate for β_2 is significantly different from zero in the pre-treatment sample, then there might be misspecification issues in A.3, and by similarity in A.2 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. The second one amounts to bias caused by measurement error in age 7 maths score.

Appendix B

Appendix to Chapter 3

B.1 Appendix figures

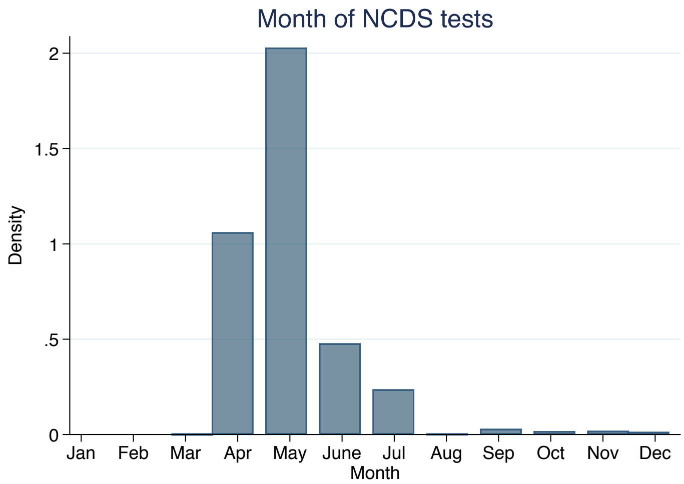


Figure B1: Histogram picturing children taking the NCDS ability tests by month (N=3448).

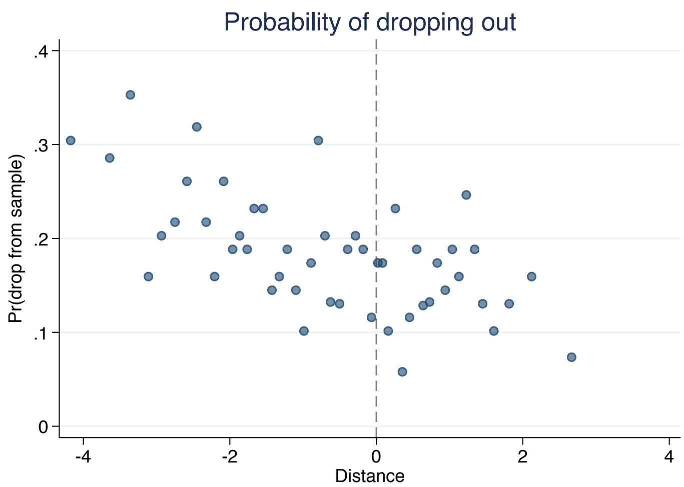


Figure B2: The probability of dropping out of the sample is a smooth function of the distance variable around the threshold (50 bins).

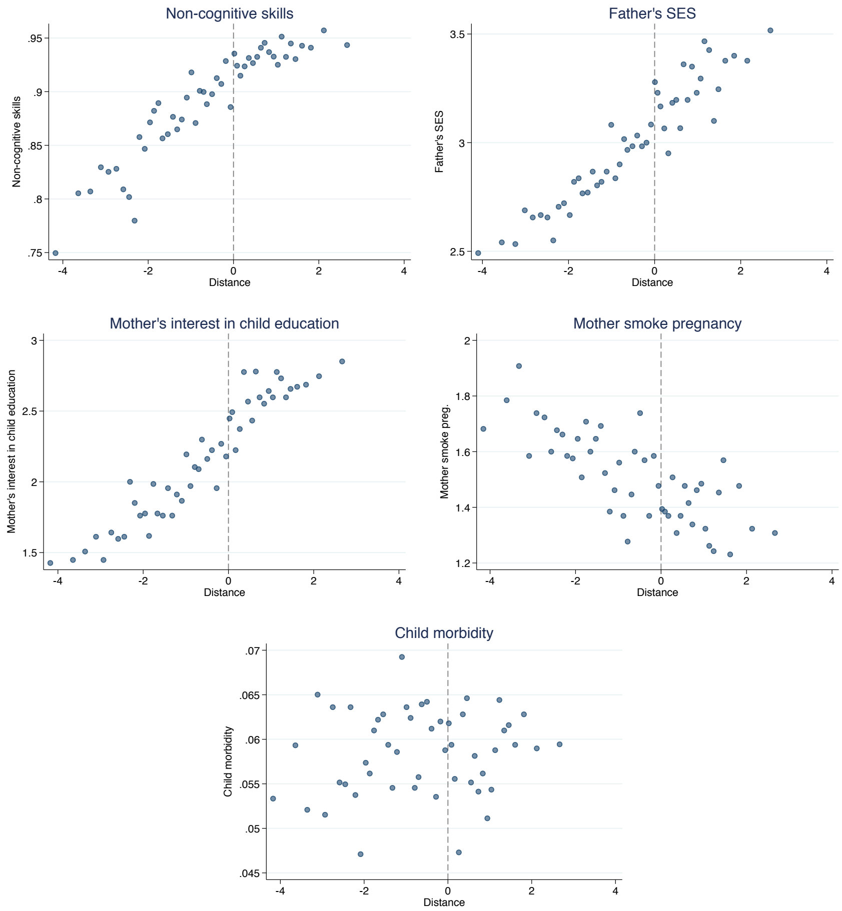


Figure B3: The plotted covariates are smooth functions of the distance variable around the threshold (50 bins).

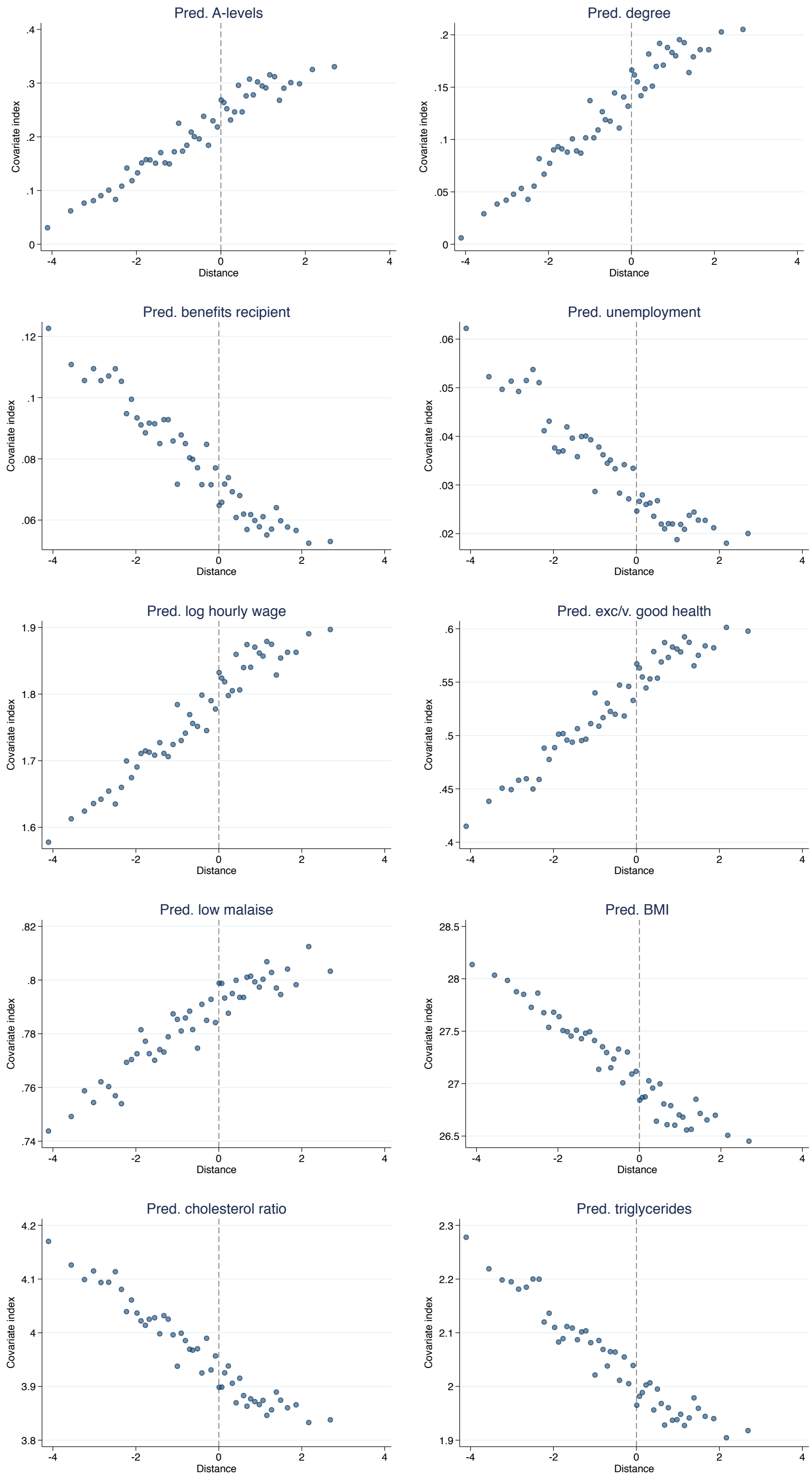
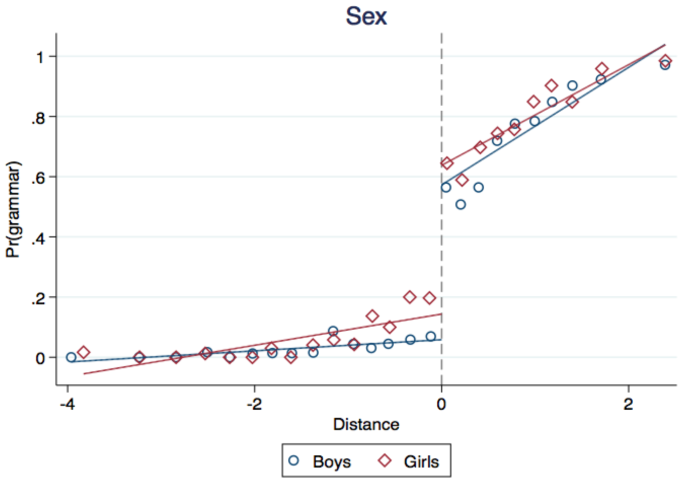
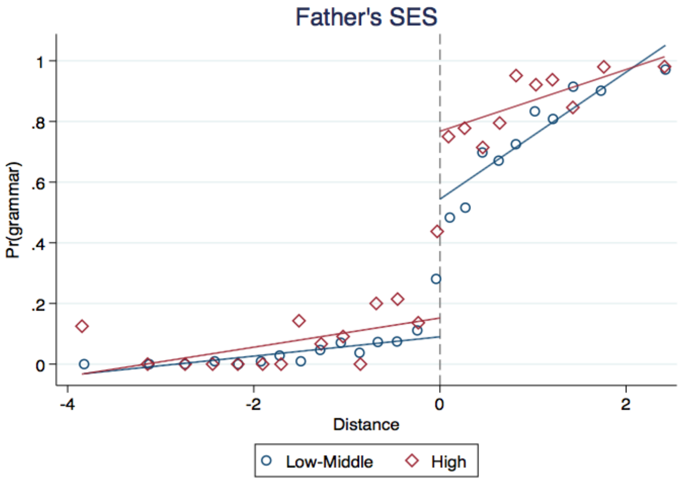


Figure B4: The covariate indices for predicted outcomes are smooth functions of the distance variable around the threshold (50 bins).





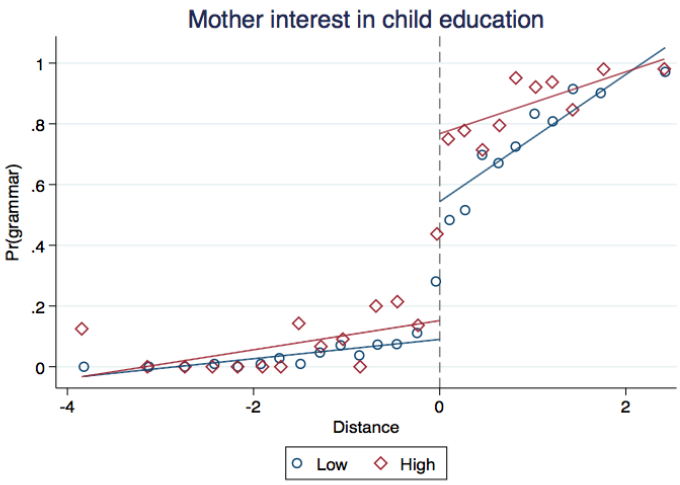


Figure B5: Probability of attending grammar as a function of distance from the cut-off, by sex, father's SES and mother's interest in child education (25 bins for each category).

B.2 Appendix tables

Table B1: Summary of related literature.

	Country	Outcomes Educ.	Labour	Health	Method
Abdulkadiroglu et al. (2014)	United States	x			RDD
Atkinson et al. (2006)	England	x			Matching and logistic regressions
Basu et al. (2018)	England			x	IV
Bonhomme and Sauder (2011)	England	x			VAR and DID
Burgess et al. (2017)*	England	x			Matching and regressions
Burgess et al. (2019)	England		x		Matching and regressions
Clark (2010)	England	x			RDD
Del Bono and Clark (2016)	Scotland	x	x	Fertility	RDD
Dobbie and Fryer (2014)	United States	x			RDD
Dustmann et al. (2017)	Germany	x	x		RDD
Galindo-Rueda and Vignoles (2005)*	England	x			VAR and IV
Guyon et al. (2012)	Northern Ireland	x			RDD
Hall (2012)	Sweden	x	x		IV
Harmon and Walker (2000)	England		x		OLS and IV
Jerrim and Sims (2018)	England	NCS			Matching
Jerrim and Sims (2019)	England	NCS			Matching and DID
Jones et al. (2011)	England			x	Matching and regressions
Jones et al. (2012)	England			x	FSD and distributional regressions
Kerckhoff (1986)	England	x			OLS regressions
Kerr et al. (2013)	Finland	Skills			DID
Kirabo Jackson (2010)	Trinidad and Tobago	x			RDD
Manning and Pischke (2006)*	England	x			VAR, IV and placebo tests
Maurin and McNally (2009)*	England	x			Regressions with pre- and post-reform cohort
Pop-Eleches and Urquiola (2013)	Romania	x			RDD

Unpublished working papers (*). Legend: Non-cognitive skills (NCS), Value-added regressions (VAR), instrumental variables (IV), difference-in-differences (DID), first-order stochastic dominance (FSD).

Table B2: Descriptive statistics of covariates for all individuals attending grammar or secondary modern for whom we have age 11 ability test scores, by whether they subsequently drop out from the sample.

	All	Dropped	Observed
Cognitive ability index	0.27	-0.06	0.34
Non-cognitive skills	0.89	0.87	0.90
Father's SES	3.00	2.81	3.01
Mother's interest	2.16	1.95	2.20
Mother smoke preg.	1.51	1.55	1.50
Child morbidity	0.06	0.03	0.06
Observations	3448	632	2816

Mother interest in child education is on a scale from 1-Little interest to 4-Over concerned. Father's SES is on a scale from 1-Low to 5-High. Maternal smoking during pregnancy is on a scale from 1-Non-smoker to 4-Heavy smoker.

Table B3: Probability of being observed in the age 16 survey as a function of ability scores and other covariates, estimated via probit models on the total sample.

	Probability: observed at age 16	
Cognitive ability index	0.0332***	0.0157
	(0.00877)	(0.0104)
Non-cognitive skills		0.442***
		(0.118)
Father's SES		0.0364*
		(0.0166)
Mother's interest		-0.000302
		(0.0144)
Mother smoke preg.		0.0354*
		(0.0158)
Child morbidity		0.430
		(0.404)
Constant	0.721***	0.144
	(0.0137)	(0.123)
Observations	10151	10151

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B4: Probability of dropping out from the sample and the LEA-specific threshold.

	Probability(drop out)
$1[A \geq c_{LEA}]$	0.0484
	(0.0929)
Distance	-0.121***
	(0.0286)
Distance $\times 1[A \geq c_{LEA}]$	0.0709
	(0.0671)
Constant	-1.035***
	(0.0596)
Observations	3448

Standard errors clustered at LEA level in parentheses. All controls included. *** $p < 0.001$

Table B5: Descriptive statistics of childhood characteristics and outcomes by type of secondary school, contrasting whole sample to compliers (selected by MSE-optimal bandwidths for A-levels).

	Grammar				**Secondary modern**			
	All		Compliers		All		Compliers	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
Cognitive skills	1.81	0.84	1.58	0.68	-0.35	1.20	0.40	0.79
Non-cognitive skills	0.94	0.08	0.94	0.08	0.88	0.12	0.91	0.09
Female	0.55	0.50	0.57	0.49	0.52	0.50	0.53	0.50
Mother's interest	2.71	0.76	2.70	0.74	1.92	1.04	2.08	1.03
Father's SES	3.33	0.91	3.30	0.90	2.85	0.81	2.95	0.81
Mother smoke preg.	1.37	0.77	1.37	0.76	1.55	0.90	1.46	0.84
Child morbidity	0.06	0.03	0.06	0.03	0.06	0.04	0.07	0.03
Observations	1091		723		1172		504	

Mother interest in child education is on a scale from 1-Little interest to 4-Over concerned. Father's SES is on a scale from 1-Low to 5-High. Maternal smoking during pregnancy is on a scale from 1-Non-smoker to 4-Heavy smoker.

Table B6: First-stage regressions with pre-selected bandwidth for each outcome.

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

Bootstrapped standard errors clustered at LEA level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B7: Human capital outcomes: reduced form regressions with automatically selected bandwidth.

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

Bootstrapped standard errors clustered at LEA level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B8: Health outcomes: reduced form regressions with automatically selected bandwidth.

	(6) High SAH	(7) Low malaise	(8) BMI	(9) Chol	(10) Trig
$1[A \geq c_{LEA}]$	0.0213	-0.0346	-0.269	-0.176	-0.352
	(0.0698)	(0.0502)	(0.686)	(0.168)	(0.217)
Distance	0.0565	0.0334	1.445	0.188	0.712*
	(0.0980)	(0.0561)	(1.037)	(0.249)	(0.332)
Distance $\times 1[A \geq c_{LEA}]$	0.00388	-0.0665	-2.735*	-0.387	-0.991*
	(0.133)	(0.0754)	(1.375)	(0.333)	(0.442)
Non-cognitive skills	0.212	0.0637	-1.432	-0.176	-0.556
	(0.203)	(0.151)	(2.015)	(0.504)	(0.656)
Female	0.0415	-0.117***	-0.868*	-0.993***	-0.973***
	(0.0364)	(0.0260)	(0.357)	(0.0881)	(0.114)
Mother's interest	0.0322	0.00416	-0.279	-0.0275	0.0887
	(0.0206)	(0.0148)	(0.198)	(0.0488)	(0.0633)
Father's SES	0.0677**	0.0106	-0.917***	-0.0981	-0.146*
	(0.0213)	(0.0153)	(0.212)	(0.0526)	(0.0682)
Mother smoke preg.	-0.0151	-0.0240	0520	-0.0741	0.00420
	(0.0219)	(0.0159)	(0.217)	(0.0524)	(0.0677)
Child morbidity	-0.218	0.330	6.785	0.514	0.442
	(0.523)	(0.381)	(5.129)	(1.281)	(1.658)
Constant	0.0794	0.818***	32.56***	5.282***	3.686***
	(0.201)	(0.149)	(2.031)	(0.505)	(0.660)
F statistic	3.580	3.060	4.678	17.263	10.599
Tot obs. available	1869	1865	1786	1498	1500
Obs. in bandwidth	771	974	716	605	592
Bandwidth	0.913	1.156	0.865	0.875	0.859

Bootstrapped standard errors clustered at LEA level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B9: All outcomes: bias-corrected model with placebo threshold at 0.2 and pre-selected bandwidth.

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

Bootstrapped standard errors clustered at LEA level in parentheses. Covariates included and omitted from the table.

Table B10: All outcomes: bias-corrected model with donut exclusion around the threshold and pre-selected bandwidth.

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

Bootstrapped standard errors clustered at LEA level in parentheses. Covariates included and omitted from the table.
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B11: Mechanisms: local polynomial regressions with each channel as the treatment variable and pre-selected bandwidth.

	(1) A-levels	(2) Degree	(3) On benefits	(4) Unemployed	(5) Log wage	(6) High SAH	(7) Low malaise	(8) BMI	(9) Chol	(10) Trig
High % girls taking GCE only	0.4172^{**}	0.2478^{*}	-0.0825	0.0054	0.1661	0.0829	-0.1329	-1.1138	-0.7539	-1.4296
	(0.1331)	(0.1101)	(0.0966)	(0.0960)	(0.3989)	(0.2889)	(0.1562)	(3.0151)	(0.6853)	(1.3547)
First-stage F	41.214	28.683	33.790	15.891	8.451	17.277	21.020	11.720	8.285	8.717
Obs. in bandwidth	1599	1538	1849	1213	791	771	974	716	605	592
Total obs.	2505	2352	2816	2116	1553	1869	1865	1786	1498	1500
Bandwidth	1.565	1.605	1.658	1.330	1.141	0.913	1.156	0.865	0.875	0.859
High % girls taking CSE only	-0.8902^{*}	-0.5080	0.1305	-0.0067	-0.1845	-0.1003	0.2212	1.1938	0.7565	1.6497
	(0.4448)	(0.2914)	(0.1456)	(0.1190)	(0.4489)	(0.3489)	(0.2857)	(3.2960)	(0.6736)	(1.5036)
First-stage F	6.374	5.390	10.306	12.188	7.142	8.674	6.595	9.618	7.840	6.431
Obs. in bandwidth	1599	1538	1849	1213	791	771	974	716	605	592
Total obs.	2505	2352	2816	2116	1553	1869	1865	1786	1498	1500
Bandwidth	1.565	1.605	1.658	1.330	1.141	0.913	1.156	0.865	0.875	0.859
Single sex	0.6837^{*}	0.3487^{*}	-0.1243	0.0066	0.1292	0.0985	-0.1496	-1.2712	-0.9652	-1.9741
	(0.2661)	(0.1673)	(0.1508)	(0.1169)	(0.3158)	(0.3537)	(0.1765)	(3.4350)	(0.9081)	(2.1981)
First-stage F	11.212	14.759	11.035	9.210	12.877	10.722	14.376	8.483	5.532	5.084
Obs. in bandwidth	1599	1538	1849	1213	791	771	974	716	605	592
Total obs.	2505	2352	2816	2116	1553	1869	1865	1786	1498	1500
Bandwidth	1.565	1.605	1.658	1.330	1.141	0.913	1.156	0.865	0.875	0.859
High % teachers left	-0.9574	-0.6141	0.1795	-0.0074	-0.2080	-0.1108	0.1858	1.4043	0.8584	1.4964
	(0.5198)	(0.4043)	(0.2432)	(0.1313)	(0.5329)	(0.3888)	(0.2218)	(3.7613)	(0.7489)	(1.3664)
First-stage F	5.615	4.482	5.897	8.663	4.440	9.596	8.741	9.833	10.884	14.898
Obs. in bandwidth	1599	1538	1849	1213	791	771	974	716	605	592
Total obs.	2505	2352	2816	2116	1553	1869	1865	1786	1498	1500
Bandwidth	1.565	1.605	1.658	1.330	1.141	0.913	1.156	0.865	0.875	0.859
Teachers get career training	-4.1580	-1.7195	0.5890	-0.0954	-11.5276	3.8245	-3.3656	5.0986	5.2047	9.9846
	(6.2892)	(2.2078)	(0.8578)	(1.7259)	(276.7844)	(42.9072)	(18.4158)	(16.2308)	(10.8415)	(20.7151)
First-stage F	0.457	0.620	0.564	0.057	0.002	0.009	0.032	0.662	0.249	0.281

Obs. in bandwidth	1599	1538	1849	1213	791	771	974	716	605	592
Total obs.	2505	2352	2816	2116	1553	1869	1865	1786	1498	1500
Bandwidth	1.565	1.605	1.658	1.330	1.141	0.913	1.156	0.865	0.875	0.859
School lacks facilities	-3.7742	-1.8606	0.3244	-0.0587	-1.2464	0.8569	-30.6583	12.6112	3.8745	6.1536
	(6.6113)	(2.9943)	(0.5008)	(1.0068)	(4.4207)	(3.6628)	(1907.1377)	(62.5648)	(8.0264)	(10.3551)
First-stage F	0.316	0.406	1.471	0.105	0.117	0.076	0.000	0.059	0.268	0.397
Obs. in bandwidth	1599	1538	1849	1213	791	771	974	716	605	592
Total obs.	2505	2352	2816	2116	1553	1869	1865	1786	1498	1500
Bandwidth	1.565	1.605	1.658	1.330	1.141	0.913	1.156	0.865	0.875	0.859

Standard errors clustered at LEA level in parentheses.

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

Table B12: Mechanisms: bias-corrected model with each channel as the treatment variable and pre-selected bandwidth.

	(1) A-levels	(2) Degree	(3) On benefits	(4) Unemployed	(5) Log wage	(6) High SAH	(7) Low malaise	(8) BMI	(9) Chol	(10) Trig
Higher than median % girls taking GCE only										
Discontinuity	0.5025*	0.3884*	-0.0365	-0.2596+	0.2828	0.0714	-0.1885	-1.2041	-0.4677	-1.2134
	(0.2105)	(0.1600)	(0.1023)	(0.1429)	(0.4186)	(0.2600)	(0.1856)	(2.8831)	(0.6967)	(1.1621)
Left of c	337	378	352	438	280	320	373	325	241	338
Right of c	442	458	405	535	339	407	455	407	295	398
Higher than median % girls taking CSE only										
Discontinuity	-0.7563+	-0.6780+	0.0409	0.3354+	-0.4533	-0.1914	0.4504	0.9677	0.4133	1.6003
	(0.4153)	(0.3691)	(0.1146)	(0.2014)	(0.4572)	(0.3738)	(0.2785)	(2.9436)	(0.7854)	(1.5920)
Left of c	467	375	346	436	248	293	269	335	166	309
Right of c	574	455	402	528	312	374	348	425	224	367
Single sex school										
Discontinuity	0.6327**	0.5193*	-0.3406+	-0.0909	0.2045	0.1307	-0.3194	-1.2002	-0.5879	-1.6697
	(0.2314)	(0.2278)	(0.1824)	(0.1583)	(0.3120)	(0.3191)	(0.2359)	(3.2684)	(0.9138)	(1.7099)
Left of c	643	367	654	370	328	359	338	331	247	338
Right of c	761	450	761	428	394	444	421	417	302	398
Above median % teachers left last year										
Discontinuity	-0.5753+	-0.6258+	0.2899	0.0392	-0.8772	-0.4102	0.2209	1.7113	0.3898	1.3811
	(0.2999)	(0.3527)	(0.1835)	(0.1089)	(0.8789)	(0.3586)	(0.2604)	(3.4202)	(0.9788)	(1.2791)
Left of c	339	404	394	375	189	437	365	309	170	263
Right of c	445	489	482	432	246	533	446	395	232	324
Teachers get career training										
Discontinuity	-2.4145	-2.5510	3.0667	0.7119	16.8045	9.1015	-3.9013	5.3745	4.4776	7.8048
	(2.6392)	(3.2847)	(8.2704)	(2.1193)	(270.4987)	(202.0571)	(20.3786)	(13.6316)	(8.6578)	(12.3329)
Left of c	456	367	604	368	271	348	439	325	304	282
Right of c	562	450	700	423	335	435	541	411	364	343
School lacks facilities										
Discontinuity	14.2627	-7.9431	3.6914	0.4250	3.8620	0.8734	-5.0617	34.3283	0.3860	4.6515

	(117.9655)	(41.7723)	(17.3504)	(1.9382)	(30.5857)	(5.3048)	(26.9486)	(273.9775)	(2.5323)	(6.6731)
Left of c	420	436	479	342	291	327	322	309	218	298
Right of c	528	511	577	399	356	414	411	394	273	360

Standard errors clustered at LEA level in parentheses. Covariates included and omitted from the table.$^{+}$ $p < 0.1$, * $p < 0.05$

Appendix C

Appendix to Chapter 4

C.1 Appendix tables

Table C1: Determinants of fixed effects (behavioural outcomes).

	Dependent variable: fixed effects (predicted from Eq 4.2)					
	Smokes	Age first drink	Junk food	Tried drugs	Truant	Sedentary
Sibship size=3	0.0262*	-0.0456	-0.0295	0.0039	0.0225*	0.0877+
	(0.0111)	(0.0542)	(0.0191)	(0.0089)	(0.0112)	(0.0469)
Sibship size=4	0.0457**	0.2350***	-0.0453+	-0.0221*	-0.0013	0.1902**
	(0.0141)	(0.0697)	(0.0235)	(0.0100)	(0.0138)	(0.0595)
Mother's age at birth	0.0036*	0.0463***	-0.0030	0.0005	0.0059***	-0.0296***
	(0.0018)	(0.0062)	(0.0027)	(0.0012)	(0.0015)	(0.0064)
Mother's birth year	0.0056***	0.0423***	0.0065**	0.0019+	0.0089***	-0.0370***
	(0.0016)	(0.0045)	(0.0024)	(0.0011)	(0.0013)	(0.0055)
Mother's highest qual	0.0018	0.0109	-0.0155**	0.0030	-0.0038	-0.0752***
	(0.0036)	(0.0169)	(0.0057)	(0.0026)	(0.0035)	(0.0150)
Family income	-0.0134	-0.0586	0.0238	-0.0040	0.0041	-0.0456
	(0.0091)	(0.0504)	(0.0153)	(0.0077)	(0.0092)	(0.0388)
Not white	0.0015	0.4821***	-0.1533***	-0.0536***	0.0307*	-0.0844
	(0.0129)	(0.0654)	(0.0229)	(0.0092)	(0.0136)	(0.0568)
North East	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(.)	(.)	(.)	(.)	(.)	(.)
North West	0.0160	-1.6444***	-0.0135	0.0066	-0.0482+	1.5466***
	(0.0288)	(0.1439)	(0.0528)	(0.0183)	(0.0278)	(0.1259)
Yorks. & Humb.	0.0093	-0.0030	-0.0297	0.0274	0.0106	0.0077
	(0.0287)	(0.1466)	(0.0549)	(0.0174)	(0.0291)	(0.1303)
East Midlands	-0.9732***	1.2251***	-1.3796***	0.0300+	-1.3808***	-0.1364
	(0.0288)	(0.1564)	(0.0540)	(0.0179)	(0.0286)	(0.1275)
West Midlands	0.0381	0.1382	-0.0292	0.0546**	-0.9297***	2.0145***
	(0.0308)	(0.1570)	(0.0540)	(0.0175)	(0.0299)	(0.1306)
East of England	0.0122	0.0198	-0.1038+	0.0492*	-0.0158	-0.0479
	(0.0283)	(0.1459)	(0.0544)	(0.0193)	(0.0276)	(0.1239)
London	-0.0040	0.0390	-0.2543***	0.0690***	-0.8116***	-0.0336
	(0.0275)	(0.1466)	(0.0527)	(0.0185)	(0.0289)	(0.1253)
South East	0.0082	1.2438***	-0.1672**	0.0170	-1.0251***	-2.2308***
	(0.0266)	(0.1456)	(0.0513)	(0.0165)	(0.0270)	(0.1215)
South West	0.0161	-0.0037	-0.0933+	0.0388*	0.0541+	-0.0743
	(0.0302)	(0.1549)	(0.0559)	(0.0193)	(0.0324)	(0.1293)
Wales	-0.0157	-0.0208	-0.0375	0.0292	0.0630	-0.0751
	(0.0357)	(0.1661)	(0.0626)	(0.0241)	(0.0390)	(0.1418)
Scotland	0.0144	0.2693+	0.0345	0.0319	0.0875**	-0.1914
	(0.0320)	(0.1548)	(0.0572)	(0.0213)	(0.0337)	(0.1319)
Northern Ireland	-0.0080	0.2635	-0.0135	0.0139	0.0218	-0.1495
	(0.0323)	(0.1647)	(0.0589)	(0.0201)	(0.0332)	(0.1330)
Lone parent HH	0.0381*	-0.0687	0.0410+	0.0100	0.0737***	0.2018***
	(0.0158)	(0.0649)	(0.0223)	(0.0110)	(0.0150)	(0.0592)
Constant	-11.0194***	-84.4424***	-12.6634**	-3.7166	-17.4007***	74.3669***
	(3.2835)	(8.9775)	(4.8530)	(2.2615)	(2.6394)	(11.0991)
Observations	2449	3703	3742	3103	4172	3967

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

Table C2: Determinants of fixed effects (non-cognitive skills).

	Dependent variable: fixed effects (predicted from Eq 4.2)		
	Prosocial	Interalising	Externalising
Sibship size=3	-0.2415***	-0.1913	0.1158
	(0.0691)	(0.1245)	(0.1380)
Sibship size=4	-0.4189***	-0.1852	-0.0764
	(0.0847)	(0.1541)	(0.1673)
Mother's age at birth	0.0175+	0.0663***	0.0892***
	(0.0099)	(0.0177)	(0.0202)
Mother's birth year	0.0121	0.1012***	0.1739***
	(0.0089)	(0.0157)	(0.0174)
Mother's highest qual	0.0265	-0.0517	-0.0551
	(0.0215)	(0.0374)	(0.0429)
Family income	0.0308	-0.1258	0.0454
	(0.0599)	(0.1005)	(0.1109)
Not white	-0.5918***	0.1472	0.4237*
	(0.0806)	(0.1449)	(0.1713)
North East	0.0000	0.0000	0.0000
	(.)	(.)	(.)
North West	0.1688	-5.1397***	3.8289***
	(0.1763)	(0.3614)	(0.3804)
Yorks. & Humb.	-0.0362	-0.4056	-0.0291
	(0.1782)	(0.3589)	(0.3883)
East Midlands	-10.8667***	-4.4937***	-4.0859***
	(0.1783)	(0.3688)	(0.3863)
West Midlands	-0.1171	0.0082	0.0164
	(0.1822)	(0.3686)	(0.3849)
East of England	0.0150	0.1122	0.0882
	(0.1828)	(0.3672)	(0.3720)
London	-3.1548***	-10.5801***	-5.9223***
	(0.1748)	(0.3491)	(0.3765)
South East	-1.6719***	-7.0691***	-7.7518***
	(0.1721)	(0.3479)	(0.3605)
South West	-0.1281	0.0659	0.2161
	(0.1804)	(0.3633)	(0.3875)
Wales	-0.0574	-0.6290	-0.1730
	(0.2289)	(0.3999)	(0.4593)
Scotland	-0.0389	-0.2237	-0.1694
	(0.1920)	(0.3872)	(0.4050)
Northern Ireland	0.1738	-0.5895	0.1779
	(0.1913)	(0.3927)	(0.4128)
Lone parent HH	-0.2112*	-0.0424	0.2864+
	(0.0837)	(0.1440)	(0.1610)
Constant	-22.7506	-196.3439***	-343.4592***
	(17.7140)	(31.5203)	(34.7574)
Observations	3750	3745	3746

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

Table C3: Birth order interacted with sex and risky behaviours

	Smokes		**Age started drinking**		**Junk food**	
Female	0.0150	0.0224	-0.0035	-0.0098	-0.0321+	-0.0702*
	(0.0104)	(0.0181)	(0.0576)	(0.0963)	(0.0194)	(0.0311)
Second child	0.0302**	0.0264	-0.1016+	-0.2937**	0.0734***	0.0219
	(0.0107)	(0.0217)	(0.0577)	(0.1000)	(0.0197)	(0.0345)
Third child	0.0402*	0.0158	-0.2770**	-0.5237**	0.0842**	-0.0167
	(0.0161)	(0.0369)	(0.0864)	(0.1674)	(0.0286)	(0.0558)
Fourth child	0.0439	-0.0029	-0.4046**	-0.9470***	0.0502	-0.0359
	(0.0272)	(0.0599)	(0.1309)	(0.2480)	(0.0468)	(0.0869)
Second × Female	-0.0234	-0.0267	-0.0197	0.0838	-0.0252	0.0092
	(0.0152)	(0.0229)	(0.0788)	(0.1261)	(0.0275)	(0.0419)
Third × Female	-0.0195	-0.0265	0.1404	0.0991	0.0293	0.1127*
	(0.0205)	(0.0316)	(0.1065)	(0.1685)	(0.0367)	(0.0537)
Fourth × Female	-0.0259	-0.0322	0.0824	0.2720	0.1268*	0.1691+
	(0.0345)	(0.0657)	(0.1687)	(0.3040)	(0.0614)	(0.0934)
Fixed-effects	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4077	4077	6123	6123	6204	6204
	Tried drugs		**Play truant**		**Sedentary behaviour**	
Female	-0.0115	-0.0102	-0.0019	-0.0010	0.4855***	0.5139***
	(0.0074)	(0.0135)	(0.0112)	(0.0185)	(0.0501)	(0.0799)
Second child	0.0197*	0.0402*	0.0229*	0.0370+	-0.0216	-0.1336
	(0.0089)	(0.0159)	(0.0116)	(0.0212)	(0.0511)	(0.0857)
Third child	0.0228+	0.0514*	0.0278	0.0785*	0.0754	-0.1101
	(0.0127)	(0.0232)	(0.0179)	(0.0344)	(0.0746)	(0.1472)
Fourth child	0.0115	0.0821*	0.0172	0.1270*	-0.1383	-0.2605
	(0.0203)	(0.0381)	(0.0283)	(0.0520)	(0.1223)	(0.2275)
Second × Female	-0.0041	-0.0168	-0.0026	0.0119	0.0060	-0.0546
	(0.0116)	(0.0184)	(0.0161)	(0.0256)	(0.0688)	(0.1035)
Third × Female	-0.0027	-0.0006	-0.0112	-0.0368	-0.0140	-0.1940
	(0.0159)	(0.0239)	(0.0222)	(0.0336)	(0.0933)	(0.1426)
Fourth × Female	0.0121	-0.0356	-0.0433	-0.1087*	0.1714	0.0176
	(0.0256)	(0.0473)	(0.0343)	(0.0526)	(0.1545)	(0.2433)
Fixed effects	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5188	5188	6930	6930	6596	6596

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

Table C4: Birth order interacted with mother's education and risky behaviours

	Smokes		**Age started drinking**		**Junk food**	
Mother has a degree	0.0197+	0.0000	0.0149	0.0000	-0.0685**	0.0000
	(0.0117)	(.)	(0.0671)	(.)	(0.0227)	(.)
Second child	0.0233*	0.0113	-0.1368**	-0.2953***	0.0547***	0.0156
	(0.0091)	(0.0207)	(0.0464)	(0.0818)	(0.0158)	(0.0287)
Third child	0.0355*	-0.0041	-0.2114**	-0.4748**	0.1107***	0.0436
	(0.0138)	(0.0367)	(0.0727)	(0.1576)	(0.0239)	(0.0534)
Fourth child	0.0407+	-0.0298	-0.3841***	-0.7979**	0.1144**	0.0493
	(0.0234)	(0.0627)	(0.1149)	(0.2463)	(0.0400)	(0.0804)
Second × Mother has a degree	-0.0178	0.0057	0.1170	0.1349	0.0259	0.0340
	(0.0156)	(0.0206)	(0.0853)	(0.1088)	(0.0293)	(0.0356)
Third × Mother has a degree	-0.0211	0.0269	0.0473	-0.0643	-0.0585	-0.0491
	(0.0231)	(0.0355)	(0.1305)	(0.1824)	(0.0440)	(0.0634)
Fourth × Mother has a degree	-0.0533*	0.0683	0.1367	-0.3384	-0.0271	-0.1243
	(0.0257)	(0.0487)	(0.1659)	(0.3204)	(0.0851)	(0.1344)
Fixed effects	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4077	4077	6123	6123	6204	6204
	Tried drugs		**Play truant**		**Sedentary behaviour**	
Mother has a degree	0.0077	0.0000	-0.0163	0.0000	-0.1500*	0.0000
	(0.0087)	(.)	(0.0120)	(.)	(0.0582)	(.)
Second child	0.0178**	0.0376**	0.0159	0.0350*	-0.0045	-0.1606*
	(0.0068)	(0.0132)	(0.0098)	(0.0179)	(0.0409)	(0.0724)
Third child	0.0223*	0.0563*	0.0240	0.0594+	0.1343*	-0.1403
	(0.0099)	(0.0231)	(0.0150)	(0.0314)	(0.0630)	(0.1429)
Fourth child	0.0192	0.0706+	-0.0006	0.0736	-0.0431	-0.1767
	(0.0172)	(0.0406)	(0.0229)	(0.0478)	(0.1006)	(0.2144)
Second × Mother has a degree	-0.0003	-0.0185	0.0274	0.0305	-0.0231	0.0178
	(0.0134)	(0.0161)	(0.0174)	(0.0200)	(0.0724)	(0.0854)
Third × Mother has a degree	-0.0049	-0.0137	0.0003	0.0022	-0.2170+	-0.2612+
	(0.0190)	(0.0264)	(0.0253)	(0.0342)	(0.1107)	(0.1510)
Fourth × Mother has a degree	-0.0110	-0.0038	-0.0047	0.0423	0.1147	-0.3044
	(0.0325)	(0.0439)	(0.0413)	(0.0656)	(0.1924)	(0.2862)
Fixed effects	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5571	5571	7459	7459	4484	4484

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

Table C5: Birth order interacted with family income and risky behaviours

	Smokes		**Age started drinking**		**Junk food**	
Family income > median	-0.0071	0.0000	0.0438	0.0000	0.0009	0.0000
	(0.0109)	(.)	(0.0627)	(.)	(0.0213)	(.)
Second child	0.0160	0.0006	-0.0987+	-0.2793**	0.0404*	0.0179
	(0.0117)	(0.0220)	(0.0560)	(0.0922)	(0.0192)	(0.0317)
Third child	0.0436*	-0.0000	-0.1758*	-0.5434**	0.0950***	0.0367
	(0.0174)	(0.0387)	(0.0864)	(0.1689)	(0.0286)	(0.0581)
Fourth child	0.0419	-0.0574	-0.2342+	-0.8702**	0.0920+	-0.0156
	(0.0277)	(0.0670)	(0.1382)	(0.2735)	(0.0477)	(0.0897)
Second × Fam. inc. > med.	0.0047	0.0280	-0.0265	0.0413	0.0398	0.0106
	(0.0147)	(0.0183)	(0.0756)	(0.0980)	(0.0258)	(0.0316)
Third × Fam. inc. > med.	-0.0279	0.0111	-0.0672	0.1168	0.0060	-0.0122
	(0.0206)	(0.0333)	(0.1070)	(0.1602)	(0.0364)	(0.0531)
Fourth × Fam. inc. > med.	-0.0226	0.1120	-0.2894+	0.0314	0.0353	0.1017
	(0.0337)	(0.0705)	(0.1682)	(0.2901)	(0.0618)	(0.0966)
Fixed-effects	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4077	4077	6123	6123	6204	6204
	Tried drugs		**Play truant**		**Sedentary behaviour**	
Family income > median	-0.0088	0.0000	-0.0136	0.0000	-0.0941+	0.0000
	(0.0082)	(.)	(0.0123)	(.)	(0.0553)	(.)
Second child	0.0112	0.0264+	0.0212+	0.0291	-0.0369	-0.1335
	(0.0086)	(0.0144)	(0.0124)	(0.0206)	(0.0512)	(0.0820)
Third child	0.0111	0.0428+	0.0273	0.0649+	0.0827	-0.0877
	(0.0117)	(0.0241)	(0.0181)	(0.0347)	(0.0746)	(0.1534)
Fourth child	0.0138	0.0448	-0.0166	0.0864	-0.1353	0.0092
	(0.0201)	(0.0502)	(0.0264)	(0.0547)	(0.1194)	(0.2294)
Second × Fam. inc. > med.	0.0124	0.0091	0.0008	0.0270	0.0332	-0.0388
	(0.0116)	(0.0144)	(0.0159)	(0.0193)	(0.0671)	(0.0818)
Third × Fam. inc. > med.	0.0208	0.0138	-0.0105	-0.0088	-0.0345	-0.2182
	(0.0160)	(0.0220)	(0.0221)	(0.0325)	(0.0943)	(0.1429)
Fourth × Fam. inc. > med.	0.0054	0.0421	0.0306	-0.0128	0.1747	-0.5303*
	(0.0258)	(0.0450)	(0.0347)	(0.0561)	(0.1538)	(0.2468)
Fixed effects	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5188	5188	6930	6930	6596	6596

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

Table C6: Birth order interacted with ethnicity and risky behaviours

	Smokes		**Age started drinking**		**Junk food**	
Not white	-0.0231+	0.0000	0.2922***	0.0000	-0.0349	0.0000
	(0.0121)	(.)	(0.0710)	(.)	(0.0238)	(.)
Second child	0.0217*	0.0090	-0.1259**	-0.2241**	0.0732***	0.0267
	(0.0090)	(0.0201)	(0.0460)	(0.0800)	(0.0157)	(0.0285)
Third child	0.0361*	-0.0150	-0.2057**	-0.3447*	0.1166***	0.0421
	(0.0149)	(0.0355)	(0.0753)	(0.1503)	(0.0248)	(0.0533)
Fourth child	0.0224	-0.0401	-0.4165***	-0.7178**	0.0867*	0.0324
	(0.0243)	(0.0710)	(0.1222)	(0.2493)	(0.0426)	(0.0852)
Second × Not white	-0.0120	0.0174	0.0605	-0.1264	-0.0510+	-0.0081
	(0.0160)	(0.0194)	(0.0878)	(0.1159)	(0.0299)	(0.0376)
Third × Not white	-0.0165	0.0571+	-0.0037	-0.4473*	-0.0665+	-0.0294
	(0.0210)	(0.0339)	(0.1173)	(0.1788)	(0.0403)	(0.0580)
Fourth × Not white	0.0269	0.0659	0.1495	-0.3896	0.0611	-0.0061
	(0.0371)	(0.0648)	(0.1790)	(0.2876)	(0.0672)	(0.1003)
Fixed effects	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4077	4077	6123	6123	6204	6204
	Tried drugs		**Play truant**		**Sedentary behaviour**	
Not white	-0.0131	0.0000	-0.0105	0.0000	0.0031	0.0000
	(0.0090)	(.)	(0.0138)	(.)	(0.0628)	(.)
Second child	0.0227**	0.0352**	0.0233*	0.0506**	-0.0089	-0.1673*
	(0.0071)	(0.0136)	(0.0096)	(0.0172)	(0.0400)	(0.0708)
Third child	0.0248*	0.0395+	0.0288+	0.0673*	0.1243+	-0.1484
	(0.0107)	(0.0227)	(0.0155)	(0.0316)	(0.0663)	(0.1429)
Fourth child	0.0222	0.0884*	-0.0154	0.0784	-0.1281	-0.3435
	(0.0198)	(0.0410)	(0.0235)	(0.0481)	(0.1052)	(0.2216)
Second × Not white	-0.0217+	-0.0179	-0.0061	-0.0264	-0.0386	0.0483
	(0.0120)	(0.0152)	(0.0184)	(0.0222)	(0.0775)	(0.0947)
Third × Not white	-0.0138	0.0305	-0.0219	-0.0151	-0.1962+	-0.1426
	(0.0159)	(0.0227)	(0.0243)	(0.0351)	(0.1019)	(0.1473)
Fourth × Not white	-0.0189	-0.0783+	0.0351	0.0128	0.2135	0.3231
	(0.0232)	(0.0475)	(0.0387)	(0.0615)	(0.1690)	(0.2471)
Fixed effects	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5188	5188	6930	6930	6596	6596

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

Table C7: Birth order interacted with sex and non-cognitive skills.

	SDQ Prosocial		SDQ Interalising		SDQ Externalising	
Female	0.8267***	0.9395***	1.2322***	1.4190***	-0.6518***	-1.0250***
	(0.0703)	(0.1179)	(0.1273)	(0.1991)	(0.1406)	(0.2257)
Second child	-0.1568*	-0.1168	0.1795	0.4679*	0.5950***	0.9043***
	(0.0742)	(0.1291)	(0.1249)	(0.2156)	(0.1440)	(0.2558)
Third child	-0.0094	-0.0112	0.0220	0.5322	0.4640*	1.0362*
	(0.1131)	(0.2256)	(0.1781)	(0.3740)	(0.2180)	(0.4271)
Fourth child	0.1491	-0.1514	0.1674	0.8134	0.4874	1.5841*
	(0.1821)	(0.3643)	(0.2836)	(0.6033)	(0.3422)	(0.6650)
Second × Female	0.0390	-0.0220	-0.3234+	-0.6800*	-0.2042	-0.2957
	(0.0989)	(0.1559)	(0.1822)	(0.2749)	(0.1990)	(0.3129)
Third × Female	-0.0300	0.0229	0.0195	-0.1690	0.2044	0.2056
	(0.1367)	(0.2119)	(0.2401)	(0.3828)	(0.2728)	(0.4148)
Fourth × Female	0.2314	0.3173	-0.5697	-1.1428+	-0.0128	0.1116
	(0.2105)	(0.3545)	(0.3983)	(0.6786)	(0.4108)	(0.6314)
Fixed effects	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6216	6216	6209	6209	6211	6211

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

Table C8: Birth order interacted with mother's education and non-cognitive skills.

	SDQ Prosocial		SDQ Interalising		SDQ Externalising	
Mother has a degree	0.2259**	0.0000	-0.1916	0.0000	-0.2039	0.0000
	(0.0819)	(.)	(0.1532)	(.)	(0.1661)	(.)
Second child	-0.0797	-0.0755	0.0128	0.1600	0.4852***	0.7502***
	(0.0586)	(0.1099)	(0.1021)	(0.1897)	(0.1135)	(0.2055)
Third child	0.0166	0.0924	-0.0170	0.4212	0.5835**	1.0522**
	(0.0898)	(0.2131)	(0.1575)	(0.3759)	(0.1779)	(0.3843)
Fourth child	0.2223	-0.0101	-0.0927	0.4600	0.5176+	1.5976**
	(0.1440)	(0.3354)	(0.2531)	(0.5841)	(0.2762)	(0.5820)
Second × Mother has a degree	-0.2581*	-0.1821	0.0555	-0.1112	0.0327	-0.0048
	(0.1106)	(0.1364)	(0.1945)	(0.2285)	(0.2154)	(0.2641)
Third × Mother has a degree	-0.2548	-0.3672	0.3796	0.1769	-0.0687	0.2976
	(0.1635)	(0.2310)	(0.2983)	(0.4243)	(0.3366)	(0.4702)
Fourth × Mother has a degree	0.2379	0.3128	0.0506	-1.2331	-0.2565	-0.1790
	(0.2810)	(0.3893)	(0.5728)	(0.8406)	(0.5725)	(1.0726)
Fixed effects	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6216	6216	6209	6209	6211	6211

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

Table C9: Birth order interacted with family income and non-cognitive skills.

	SDQ Prosocial		**SDQ Interalising**		**SDQ Externalising**	
Family income > median	-0.0772	0.0000	-0.4351**	0.0000	0.0671	0.0000
	(0.0762)	(.)	(0.1383)	(.)	(0.1544)	(.)
Second child	-0.2260**	-0.2169+	-0.0463	0.0382	0.5207***	0.7893***
	(0.0738)	(0.1243)	(0.1292)	(0.2166)	(0.1404)	(0.2303)
Third child	-0.0692	-0.1357	-0.1111	0.3764	0.6478**	1.2637**
	(0.1103)	(0.2341)	(0.1884)	(0.4140)	(0.2165)	(0.4235)
Fourth child	0.1570	-0.3911	-0.2488	0.1920	0.6525*	1.6195*
	(0.1738)	(0.3823)	(0.2958)	(0.6574)	(0.3273)	(0.6289)
Second × Fam. inc. > med.	0.1725+	0.1445	0.1236	0.1676	-0.0527	-0.0623
	(0.0971)	(0.1203)	(0.1686)	(0.2043)	(0.1875)	(0.2314)
Third × Fam. inc. > med.	0.0818	0.2143	0.3002	0.1276	-0.1613	-0.2699
	(0.1352)	(0.2101)	(0.2393)	(0.3715)	(0.2684)	(0.4014)
Fourth × Fam. inc. > med.	0.2038	0.7537*	0.3105	0.2138	-0.3658	0.0258
	(0.2129)	(0.3681)	(0.3965)	(0.6866)	(0.4133)	(0.7107)
Fixed effects	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6216	6216	6209	6209	6211	6211

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

Table C10: Birth order interacted with ethnicity and non-cognitive skills.

	SDQ Prosocial		**SDQ Interalising**		**SDQ Externalising**	
Not white	0.1963*	0.0000	-0.5701***	0.0000	-0.7771***	0.0000
	(0.0875)	(.)	(0.1443)	(.)	(0.1703)	(.)
Second child	-0.1084+	-0.0939	-0.0378	0.0650	0.5237***	0.8102***
	(0.0584)	(0.1086)	(0.1043)	(0.1908)	(0.1153)	(0.2064)
Third child	-0.0419	-0.0076	0.1177	0.4113	0.5029**	1.1447**
	(0.0924)	(0.2031)	(0.1638)	(0.3693)	(0.1855)	(0.3919)
Fourth child	0.3936**	0.0152	-0.1181	0.0895	0.4615	1.7399**
	(0.1511)	(0.3340)	(0.2789)	(0.6032)	(0.3005)	(0.6394)
Second × Not white	-0.1221	-0.1469	0.2360	0.3007	-0.1297	-0.2497
	(0.1141)	(0.1430)	(0.1833)	(0.2290)	(0.2095)	(0.2594)
Third × Not white	0.0493	-0.0034	-0.2797	0.1997	0.2241	-0.1045
	(0.1527)	(0.2324)	(0.2597)	(0.4081)	(0.2997)	(0.4189)
Fourth × Not white	-0.4322+	-0.1310	0.0531	0.7377	0.0603	-0.4431
	(0.2311)	(0.3990)	(0.4166)	(0.7227)	(0.4554)	(0.6592)
Female	0.8478***	0.9476***	1.0798***	1.0514***	-0.7013***	-1.1153***
	(0.0459)	(0.0740)	(0.0805)	(0.1225)	(0.0894)	(0.1380)
Fixed effects	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6216	6216	6209	6209	6211	6211

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

Table C11: Birth order and parental investments.

	Breastfed	**Months bfed**	**Int. school**	**Help hmk**	**Feels supp.**	**Talk to par.**
Panel A						
Second child	-0.0136	-0.3232	-0.0681**	-0.0858***	-0.0392+	-0.0380
	(0.0334)	(0.4110)	(0.0217)	(0.0232)	(0.0234)	(0.0261)
Third child	-0.0426	-0.5027	-0.1153**	-0.1022*	-0.0433	-0.0018
	(0.0679)	(0.6503)	(0.0416)	(0.0452)	(0.0453)	(0.0516)
Fourth child	-0.0448	-2.7469*	-0.1551*	-0.1852*	-0.0812	-0.0084
	(0.0883)	(1.1436)	(0.0682)	(0.0760)	(0.0680)	(0.0798)
Female	0.0248	0.2045	-0.0166	0.0210	-0.0471**	-0.0218
	(0.0253)	(0.3208)	(0.0152)	(0.0163)	(0.0163)	(0.0173)
Panel B						
Any older siblings	-0.0057	-0.0420	-0.0467*	-0.0685***	-0.0326	-0.0493*
	(0.0298)	(0.3834)	(0.0185)	(0.0197)	(0.0203)	(0.0219)
Female	0.0254	0.2264	-0.0157	0.0218	-0.0465**	-0.0221
	(0.0254)	(0.3217)	(0.0153)	(0.0164)	(0.0163)	(0.0174)
Panel C						
Older sib, diff sex	-0.0213	-0.0758	-0.0610**	-0.0636**	-0.0157	-0.0278
	(0.0325)	(0.4069)	(0.0204)	(0.0215)	(0.0218)	(0.0231)
Older sib, same sex	0.0090	-0.0092	-0.0257	-0.0756***	-0.0572*	-0.0814**
	(0.0333)	(0.4528)	(0.0207)	(0.0226)	(0.0235)	(0.0257)
Female	0.0253	0.2263	-0.0157	0.0221	-0.0467**	-0.0224
	(0.0254)	(0.3218)	(0.0152)	(0.0164)	(0.0163)	(0.0173)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2602	2551	6200	5191	6226	6055

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

Table C12: Summary of related literature

Study	Country	Outcomes	Method	Findings
Argys et al. (2006)	US	Health behaviours	OLS	Negative
Bagger et al. (2018)*	Denmark	Education	Theoretical model + FE and IV	Negative
Barclay and Kolk (2017)	Sweden	Mortality	Survival analysis	Positive (note mortality outcome)
Barclay and Myrskylä (2014)	Sweden	Physical fitness	FE	Negative.
Barclay et al. (2017)	Sweden	College major	FE	First-born more likely to apply to and graduate in medicine and engineering. Later-born more likely to study journalism, business or art. BO patterns are stronger in high SES families.
Behrman and Taubman (1986)	US	Education, earnings	Theoretical model and OLS	Negative
Bertoni and Brunello (2016)	Multi-country	Earnings	OLS, separately by family size.	Negative but short-lived effect. Later born's wages catch up in approx 10 years. May be due to lower risk aversion in later born.
Björkegren and Svaleryd (2017)*	Sweden	Health at birth, hospitalizations and mental health in adolescence	FE	Positive at birth, then reverted later.
Black et al. (2005)	Norway	Education, earnings, teenage pregnancy	IV and FE	Negative
Black et al. (2016)	Norway	Health, health behaviours	FE	Positive for biomarkers, negative for happiness and self-assessed health
Black et al. (2018)	Sweden	Non-cognitive skills	FE	Negative, stronger for men
Bonesrønning and Massih (2011)	Norway	Education	OLS	Negative. BO effects are stronger for less educated mothers.

(continued)

Study	Country	Outcomes	Method	Findings
Booth and Kee (2009)	UK	Education	OLS with purpose-built index	Negative
Breining et al. (2017)*	Denmark and Florida	Delinquency	FE	Negative, partly due to parental time investment.
Brenøe and Molitor (2018)	Denmark	Health at birth and pre-natal investments	FE	Positive effect of higher birth order on health at birth, reverted in adolescence. Mother's antenatal care higher for earlier born.
Buckles and Kolka (2014)	US	Prenatal investments	FE	Negative
Cho (2011)	Korea	Education	OLS	No effect of BO once accounting for mother's age at first birth.
Conley and Glauber (2006)	US	Education	IV	No effect of sibhip size for first-born boys. For second-born boys, sibhip size negatively correlated with education.
de Haan (2010)	US	Education	IV	Negative
de Haan et al. (2014)	Ecuador	Education, child labour	FE	Positive, partly due to mother's time investment
Emerson and Souza (2008)*	Brazil	Education, child labour	Probit models	Positive.
Grinberg (2015)	US	Occupational choice	OLS and IV	First-born more likely to be in managerial positions.
Hanushek (1992)	US	Education	OLS	Negative correlation explained by family size.
Hotz and Pantano (2015)	US	Education, parental disciplinary restrictions	FE	Negative.

(continued)

Study	Country	Outcomes	Method	Findings
Iacovou (2008)	UK	Education	OLS (dummy var)	Negative
Kumar (2016)	India	Education	IV	Positive.
Lehmann et al. (2018)	US	Cognitive/ non-cognitive skills and various	FE	Negative, largely explained by parental behaviour
Mechoulan and Wolff (2015)	France	Education, occupation and parental transfers.	Random effects and FE ordered models	Negative effects on education and occupation. Also shown parental transfers are higher for firstborns.
Pavan (2016)	US	Cognitive skills	FE	Negative, explained by parental investments
Rees et al. (2008)	US	Sport	OLS	Various
Tenikué and Tequame (2017)*	Nigeria	Teenage pregancy	OLS	Women born in families with older brothers less likely to experience a teenage pregnancy.

* Unpublished working papers at the time of writing.

C.2 Strengths and Difficulties Questionnaire

Table C13: 25-item Strengths and Difficulties Questionnaire

I try to be nice to other people. I care about their feelings
I am restless, I cannot stay still for long
I get a lot of headaches, stomach-aches or sickness
I usually share with others (food, games, pens etc.)
I get very angry and often lose my temper
I am usually on my own. I generally play alone or keep to myself
I usually do as I am told
I worry a lot
I am helpful if someone is hurt, upset or feeling ill
I am constantly fidgeting or squirming
I have one good friend or more
I fight a lot. I can make other people do what I want
I am often unhappy, down-hearted or tearful
Other people my age generally like me
I am easily distracted, I find it difficult to concentrate
I am nervous in new situations. I easily lose confidence
I am kind to younger children
I am often accused of lying or cheating
Other children or young people pick on me or bully me
I often volunteer to help others (parents, teachers, children)
I think before I do things
I take things that are not mine from home, school or elsewhere
I get on better with adults than with people my own age
I have many fears, I am easily scared
I finish the work I'm doing.

Source: Goodman et al. (1998).

C.3 Unbalanced panel analysis and discussion

In the context of this paper, the data is unbalanced by default, since families of two, three and four children are included in the sample. Different family size is not a problem for the estimates, since it is captured by the mother fixed effects. However, the estimates might still be biased, if children from certain families are missing from the survey not randomly, but according to certain characteristics. Table C14 shows significant differences in baseline characteristics of households where all children are observed in the survey (the 'balanced' households) versus those where at least one child is not observed (the 'unbalanced' households). On average, unbalanced households have children who were born more recently, they are slightly larger and siblings display a shorter distance in years between them. Their mothers are younger, both at birth and in general, although they display

Table C14: Descriptive statistics for balanced and unbalanced households.

	Balanced	Unbalanced	Difference	p-value
Female	0.49	0.50	-0.01	0.297
Birth year	1995.99	1998.81	-2.83	0.000
Not white	0.23	0.32	-0.09	0.000
Sibship size	1.60	1.98	-0.38	0.000
Years from closest older sibling	3.97	3.33	0.64	0.000
Years from eldest sibling	5.39	4.14	1.25	0.000
Mother's age at birth	29.51	26.21	3.30	0.000
Mother's birth year	1966.49	1972.60	-6.11	0.000
Mother's highest qualification	3.82	3.84	-0.01	0.704
Gross household monthly income	3967.48	3382.42	585.06	0.000
Lone parent household	0.19	0.22	-0.03	0.001
Observations	7265	2883		

Region and ethnic group not included.

similar education levels to balanced households. They also have lower monthly income and they are slightly more likely to be lone parent families.

I first conduct a test following Verbeek and Nijman (1992), adding as a regressor to the main equation a binary indicator for whether any child from a family is missing in the survey. The rationale for the test is that if children are missing at random, then the indicator should not be correlated with the outcomes, once other covariates are controlled for. Since the indicator is constant between siblings, the fixed-effect specification cannot be implemented. Moreover, given the differences observed in mean characteristics, it is not surprising that some of the indicators in Table C15, showing results of the test, are significant. The indicator for children missing is associated with a decrease in the age of first drink, and increases in the probability of trying drugs and playing truant. However, two observations are in order. First, children from unbalanced households are overall more likely to engage in risky behaviours. Second, even without observing some children from these households, I find a significant effect of birth order. Therefore, the estimates presented in the paper may actually underestimate the true birth–order effect. I thus conduct the main regressions again, but separately for balanced and unbalanced households. As shown in Table C16, fixed-effect estimates for the balanced families confirm the birth order effects found in the main specification, showing even larger associations. Conversely, for the unbalanced families, coefficients are largely insignificant. This may be due to the missing observations making it harder in practice to observe birth order effects, or to birth order effects not being as strong in families where children are already more likely to engage in risky behaviours due to other characteristics. Overall then, the unbalanced character of the panel does not invalidate the main results.

Table C15: Verbeek and Nijman (1992) style test

	(1) Smokes	(2) Age first drink	(3) Junk food	(4) Tried drugs	(5) Truant	(6) Sedentary behaviour	(7) Prosocial	(8) Interalising	(9) Externalising
Second child	0.0224*	-0.2136***	0.0621***	0.0235***	0.0325***	-0.0139	-0.1382*	0.0098	0.5138***
	(0.0088)	(0.0439)	(0.0145)	(0.0067)	(0.0090)	(0.0368)	(0.0540)	(0.0929)	(0.1041)
Third child	0.0382*	-0.4062***	0.1018***	0.0332**	0.0437**	0.0780	-0.0264	0.0179	0.6100***
	(0.0150)	(0.0735)	(0.0238)	(0.0105)	(0.0154)	(0.0616)	(0.0895)	(0.1552)	(0.1739)
Fourth child	0.0429	-0.6401***	0.1136**	0.0334	0.0258	-0.0430	0.2556	-0.1259	0.5375*
	(0.0246)	(0.1136)	(0.0398)	(0.0186)	(0.0226)	(0.0972)	(0.1411)	(0.2546)	(0.2715)
Not all children observed	0.0094	-0.3727***	0.0059	0.0198*	0.0373***	0.0216	-0.0025	-0.0334	0.0791
	(0.0106)	(0.0516)	(0.0185)	(0.0080)	(0.0110)	(0.0449)	(0.0668)	(0.1220)	(0.1325)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4076	6120	6201	5185	6927	6593	6216	6209	6211

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

Table C16: Birth order and risky behaviours, distinguishing between balanced and unbalanced panel.

	Smokes		**Age started drinking**		**Junk food**	
	Balanced	Unbalanced	Balanced	Unbalanced	Balanced	Unbalanced
Second child	-0.0004	0.0405	-0.2841**	-0.1975	0.0160	0.0289
	(0.0230)	(0.0306)	(0.0872)	(0.1346)	(0.0303)	(0.0590)
Third child	-0.0210	0.0713	-0.4647**	-0.7516*	0.0182	0.0399
	(0.0392)	(0.0639)	(0.1615)	(0.3063)	(0.0561)	(0.1209)
Fourth child	-0.0379	-0.4010	-0.8894***	0.1929	-0.0009	0.0831
	(0.0670)	(0.3455)	(0.2459)	(0.6058)	(0.0845)	(0.2272)
Fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2375	1701	4210	1910	4235	1966
	Tried drugs		**Play truant**		**Sedentary behaviour**	
	Balanced	Unbalanced	Balanced	Unbalanced	Balanced	Unbalanced
Second child	0.0348*	0.0263	0.0484*	0.0373	-0.1873*	-0.0714
	(0.0146)	(0.0242)	(0.0188)	(0.0348)	(0.0760)	(0.1451)
Third child	0.0594*	0.0383	0.0644	0.0735	-0.2843	0.1352
	(0.0251)	(0.0375)	(0.0336)	(0.0660)	(0.1508)	(0.3033)
Fourth child	0.0777	0.0640	0.0881	-0.0219	-0.3626	0.6551
	(0.0435)	(0.0707)	(0.0506)	(0.1475)	(0.2264)	(0.6404)
Fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3496	1689	4597	2330	4451	2142
	SDQ Prosocial		**SDQ Interalising**		**SDQ Externalising**	
	Balanced	Unbalanced	Balanced	Unbalanced	Balanced	Unbalanced
Second child	-0.2254*	0.2419	0.1381	0.2132	0.8275***	0.8970
	(0.1111)	(0.2504)	(0.2004)	(0.3969)	(0.2127)	(0.4673)
Third child	-0.1393	0.5187	0.4447	0.8850	1.1301**	1.6898
	(0.2109)	(0.5359)	(0.3846)	(0.8779)	(0.3980)	(0.9388)
Fourth child	-0.2753	1.3466	0.3061	0.1928	1.5085*	2.7113
	(0.3300)	(1.1788)	(0.6016)	(1.6158)	(0.6096)	(2.4590)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4245	1971	4241	1968	4241	1970

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

C.4 Alternative definitions of birth order

Based on Equations (4.1) and (4.2), I run similar models with alternative treatments. These alternative specifications define treatment as being a middle child compared to everything else; being the youngest child compared to everything else; and being a second and middle child, second and last or anything else. All models confirm the main results, that is, younger children tend to engage more in riskier behaviours and have lower non-cognitive skills, as measured by the SDQ scores. For most outcomes, not being the first child appears to drive the observed effect of birth order, and the magnitude of the effect is usually larger for the youngest child. Only for prosocial score, being a middle child (i.e. having at least one sibling before and one after) plays a significant and negative role.

C.5 Parental investment index

I construct the index P_i by principal component analysis on four binary indicators of parental investments that are highly correlated and that differ significantly by birth order: whether parents are interested in child's school, whether parents help with homework, whether the child always feels supported by the family and whether they talk to parents when upset.

C.6 Beliefs about risky behaviours

Children were asked to answer the following: How much do you think PEOPLE RISK harming themselves, physically and in other ways, if they...

1 Smoke one or more packs of cigarettes per day

2 Have five or more alcoholic drinks each weekend

3 Smoke cannabis (marijuana or hash) regularly

Possible answers were No risk, Slight risk, Moderate risk, Great risk, Don't know. I exclude children who select the latter option, and therefore final score was 0 to 4 for each dimension, where 0 denoted lowest risk aversion, and 4 highest.

Abbreviations

2SLS	Two-stage least squares
BMI	Body mass index
BSAG	Bristol Social Adjustment Guide
CIA	Conditional independence assumption
CRP	C-Reactive Protein
CSE	Certificate of Secondary Education
CCT	Calonico, Cattaneo and Titiunik (2014)
GCE	General Certificate of Education
GC	Grammar and comprehensive
IV	Instrumental variable
LATE	Local average treatment effect
LEA	Local Education Authority
MSE	Mean squared error
NCDS	National Child Development Study
OLS	Ordinary least squares
PCA	Principal component analysis
pp	Percentage points
RCT	Randomised controlled trial
RDD	Regression discontinuity design
SDQ	Strengths and difficulties questionnaire
SES	Socio-economic status
SMC	Secondary modern and comprehensive
SD	Standard deviation(s)

References

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

Adler, A. (1928). 'Characteristics of the first, second and third child'. *Children* 3 (5), pp. 14–52.

Akresh, R., Caruso, G. D., and Thirumurthy, H. (2014). 'Medium-Term Health Impacts of Shocks Experienced In Utero and After Birth: Evidence from Detailed Geographic Information on War Exposure'. NBER Working Paper No. 20763.

Allen, R. (2016a). 'Inequalities in access to teachers in selective schooling areas'. [online] Accessed 18th November 2019. URL: `https://ffteducationdatalab.org.uk/2016/06/inequalities-in-access-to-teachers-in-selective-schooling-areas/`.

— (2016b). 'When is a comprehensive school actually a secondary modern ?' [online] Accessed 17th November 2019. URL: `https://educationdatalab.org.uk/2016/11/when-is-a-comprehensive-school-actually-a-secondary-modern/`.

Almond, D. and Currie, J. (2011). 'Killing me softly: The fetal origins hypothesis'. *Journal of Economic Perspectives* 25 (3), pp. 153–172.

Altonji, J. G., Cattan, S., and Ware, I. (2017). 'Identifying sibling influence on teenage substance use'. *Journal of Human Resources* 52 (1), pp. 1–47.

American Psychological Association (2018). 'Externalizing Internalizing'. APA Dictionary of Psychology. [online] Accessed 10 October 2019. URL: `https://dictionary.apa.org/externalizing-internalizing`.

Andrews, J., Hutchinson, J., and Johnes, R. (2016). *Grammar schools and social mobility*. Education Policy Institute Report [online] Accessed on 26th November 2019. URL: `https://epi.org.uk/wp-content/uploads/2018/01/Grammar-schools-and-social-mobility%7B%5C_%7D.pdf`.

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

Angrist, J. D. and Krueger, A. B. (1991). 'Does Compulsory School Attendance Affect Schooling and Earnings?' *Quarterly Journal of Economics* 106 (4), pp. 979–1014.

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

Argys, L. M., Rees, D. I., Averett, S. L., and Witoonchart, B. (2006). 'Birth order and risky adolescent behavior'. *Economic Inquiry* 44 (2), pp. 215–233.

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

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

Bagger, J., Birchenall, J. a., Urzúa, S., and Holloway, R. (2018). 'Education, Birth Order, and Family Size'. Unpublished manuscript.

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

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

Barclay, K. J. and Kolk, M. (2017). 'The Long-Term Cognitive and Socioeconomic Consequences of Birth Intervals: A Within-Family Sibling Comparison Using Swedish Register Data'. *Demography* 54 (2), pp. 459–484.

Barclay, K., Hällsten, M., and Myrskylä, M. (2017). 'Birth Order and College Major in Sweden'. *Social Forces* 96 (2), pp. 629–660.

Barclay, K. and Myrskylä, M. (2014). 'Birth order and physical fitness in early adulthood: Evidence from Swedish military conscription data'. *Social Science and Medicine* 123, pp. 141–148.

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

Basu, A. (2014). 'Estimating person-centered treatment (PeT) effects using instrumental variables: an application to evaluating prostate cancer treatments'. *Journal of Applied Econometrics* 29 (4), pp. 671–691.

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* 57, pp. 1–14.

Becker, G. S. (1960). 'An Economic Analysis of Fertility'. *Demographic and Economic Change in Developed Countries.* Ed. by G. B. Roberts. Columbia University Press. Chap. 7, pp. 209–240.

Becker, G. S. (1962). 'Investment in Human Capital : A Theoretical Analysis'. *Journal of Political Economy* 70 (5), pp. 9–49.

Becker, G. S. and Lewis, H. G. (1973). 'On the Interaction between the Quantity and Quality of Children'. *Journal of Political Economy* 81 (2), S279–S288.

Behrman, J. R. and Taubman, P. (1986). 'Birth Order, Schooling, and Earnings'. 4 (3), pp. 121–145.

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

Bertoni, M. and Brunello, G. (2016). 'Later-borns Don't Give Up: The Temporary Effects of Birth Order on European Earnings'. *Demography* 53 (2), pp. 449–470.

Biglan, A., Brennan, P. A., Foster, S. L., and Holder, H. D. (2004). *Helping adolescents at risk: Prevention of multiple problem behaviors.* New York, NY, US: Guilford Press.

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

Björkegren, E. and Svaleryd, H. (2017). 'Birth order and child health'. *IFAU Working Paper 2017/16, Institute for Evaluation of Labour Market and Education Policy.*

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2005). 'The more the merrier? The effect of family size and birth order on children's education'. *Quarterly Journal of Economics* 120 (2), pp. 669–700.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2016). 'Healthy(?), wealthy, and wise: Birth order and adult health'. *Economics & Human Biology* 23, pp. 27–45.

Black, S. E., Grönqvist, E., and Öckert, B. (2018). 'Born to Lead? The Effect of Birth Order on Non-Cognitive Abilities'. *Review of Economics and Statistics* 100 (2), pp. 274–286.

Blake, J. (1981). 'Family Size and the Quality of Children'. *Demography* 18 (4), pp. 421–442.

Bolton, P. (2012). 'Education: Historical statistics'.

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

Bonesrønning, H. and Massih, S. S. (2011). 'Birth order effects on young students' academic achievement'. *Journal of Socio-Economics* 40 (6), pp. 824–832.

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

Booth, A. L. and Kee, H. J. (2009). 'Birth order matters: the effect of family size and birth order on educational attainment'. *Journal of Population Economics* 22 (2), pp. 367–397.

Breining, S., Doyle, J., Figlio, D., Karbownik, K., and Roth, J. (2017). 'Birth Order and Delinquency: Evidence from Denmark and Florida'. NBER Working Paper No. 23038.

Brenøe, A. A. and Molitor, R. (2018). 'Birth order and health of newborns What can we learn from Danish registry data?' *Journal of Population Economics* 31 (2), pp. 363–395.

Brinch, C., Mogstad, M., and Wiswall, M. (2017). 'Beyond LATE with a discrete instrument'. *Journal of Political Economy* 125 (4), pp. 985–1039.

Bristow, M. (2016). *The Eleven-Plus Book: Genuine Exam Questions From Yesteryear*. 3rd. Michael O'Mara Books.

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

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

Brunello, G. and Rocco, L. (2017). 'The labor market effects of academic and vocational education over the life cycle: Evidence based on a British Cohort'. *Journal of Human Capital* 11 (1), pp. 106–166.

Buckles, K. and Kolka, S. (2014). 'Prenatal investments, breastfeeding, and birth order'. *Social Science and Medicine* 118, pp. 66–70.

Bunting, B. P. and Mooney, E. (2001). 'The Effects of Practice and Coaching on Test Results for Educational Selection at Eleven Years of Age'. *Educational Psychology* 21 (3), pp. 243–253.

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.

— (2018). 'Access to grammar schools by socio-economic status'. *Environment and Planning A: Economy and Space* 50 (7), pp. 1381–1385.

Burgess, S., Dickson, M., and Macmillan, L. (2019). ‘Do selective schooling systems increase inequality?’ *Oxford Economic Papers*. gpz028, pp. 1–24.

Butcher, K. F. and Case, A. (1994). ‘The effect of sibling sex composition on women’s education and earnings’. *Quarterly Journal of Economics* 109 (3), pp. 531–563.

Byrne, D. and Williamson, W. (1976). *Effect of Local Education Authority Resources and Policies on Educational Attainment, 1972-1974*. [computer file]. Colchester, Essex: UK Data Archive SN: 199. URL: http://dx.doi.org/10.5255/UKDA-SN-199-1.

Byrnes, J. P., Miller, D. C., and Schafer, W. D. (1999). ‘Gender differences in risk taking: A meta-analysis’. *Psychological Bulletin* 125 (3), pp. 367–383.

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

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018a). ‘Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs’. *Econometrics Journal (forthcoming).*

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

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2018b). ‘Regression Discontinuity Designs Using Covariates’. *Review of Economics and Statistics* 101 (3), pp. 442–451.

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

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

Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., and Pan, Y. (2014). ‘Early Childhood Investments Substantially Boost Adult Health’. *Science* 343 (6178), pp. 1478–1485.

Campbell, R., Starkey, F., Holliday, J., Audrey, S., Bloor, M., Parry-Langdon, N., Hughes, R., and Moore, L. (2008). ‘An informal school-based peer-led intervention for smoking prevention in adolescence (ASSIST): a cluster randomised trial’. *The Lancet* 371 (9624), pp. 1595–1602.

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

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

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

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

Cawley, J., Conneely, K., Heckman, J., and Vytlacil, E. (1997). 'Cognitive Ability, Wages, and Meritocracy'. *Intelligence, Genes, and Success: Scientists Respond to The Bell Curve.* Ed. by B. Devlin, S. E. Fienberg, D. P. Resnick, and K. Roeder. New York, NY: Springer New York. Chap. 8, pp. 179–192.

Charness, G. and Gneezy, U. (2012). 'Strong Evidence for Gender Differences in Risk Taking'. *Journal of Economic Behavior and Organization* 83 (1), pp. 50–58.

Cho, H. (2011). 'Birth order and education: Evidence from a Korean cohort'. *Economics Letters* 110 (3), pp. 200–202.

Clark, B. D. and Royer, H. (2013). 'The Effect of Education on Adult Mortality and Health: Evidence from Britain'. *American Economic Review* 103 (6), pp. 2087–2120.

Clark, D. (2010). 'Selective Schools and Academic Achievement'. *The B.E. Journal of Economic Analysis & Policy* 10 (1), pp. 1–38.

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

Conley, D. and Glauber, R. (2006). 'Parental Educational Investment and Children's Academic Risk. Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility'. *Journal of Human Resources* 41 (4), pp. 722–737.

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

Cribb, J., Jesson, D., Sibieta, L., Skipp, A., and Vignoles, A. (2013). 'Poor Grammar. Entry into Grammar Schools for disadvantaged pupils in England.' The Sutton Trust Report [online] Accessed on 25th November 2019. URL: `http://dera.ioe.ac.uk/id/eprint/30279`.

Cruz-Gonzalez, M., Fernández-Val, I., and Weidner, M. (2017). 'Bias corrections for probit and logit models with two-way fixed effects'. *The Stata Journal* 17 (3), pp. 517–545.

Cunha, B. Y. F., Heckman, J. J., and Schennach, S. M. (2010). 'Estimating the Technology of Cognitive and Noncognitive Skill Formation'. *Econometrica* 78 (3), pp. 883–931.

Currie, J. and Almond, D. (2011). 'Human capital development before age five'. *Handbook of Labor Economics*. Ed. by O. C. Ashenfelter and D. Card. Vol. 4. (Part B). Elsevier. Chap. 15, pp. 1315–1486.

de Haan, M. (2010). 'Birth order, family size and educational attainment'. *Economics of Education Review* 29 (4), pp. 576–588.

de Haan, M., Plug, E., and Rosero, J. (2014). 'Birth Order and Human Capital Development Evidence from Ecuador'. *Journal of Human Resources* 49 (2), pp. 359–391.

Dearden, L., Ferri, J., and Meghir, C. (2002). 'The Effect of School Quality on Educational Attainment and Wages'. *Review of Economics and Statistics* 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* 8 (1), pp. 150–176.

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

Dong, Y. (2018). 'Jump or Kink? Regression Probability Jump and Kink Design for Treatment Effect Evaluation'. Unpublished manuscript.

Dustmann, C., Puhani, P. A., and Schönberg, U. (2017). 'The Long-term Effects of Early Track Choice'. *Economic Journal* 127 (603), pp. 1348–1380.

Emerson, P. M. and Souza, A. P. (2008). 'Birth Order, Child Labor, and School Attendance in Brazil'. *World Development* 36 (9), pp. 1647–1664.

Eriksson, I., Undén, A.-L., and Elofsson, S. (2001). 'Self-rated health. Comparisons between three different measures. Results from a population study'. *International Journal of Epidemiology* 30 (2), pp. 326–333.

Evensen, M., Lyngstad, T. H., Melkevik, O., and Mykletun, A. (2016). 'The Role of Internalizing and Externalizing Problems in Adolescence for Adult Educational Attainment: Evidence from Sibling Comparisons using Data from the Young HUNT Study'. *European Sociological Review* 32 (5), pp. 552–566.

Fernández-Val, I. and Weidner, M. (2016). 'Individual and time effects in nonlinear panel models with large N, T'. *Journal of Econometrics* 192 (1), pp. 291–312.

Foley, B. C., Shrewsbury, V. A., Hardy, L. L., Flood, V. M., Byth, K., and Shah, S. (2017). 'Evaluation of a peer education program on student leaders' energy balance-related behaviors'. *BMC Public Health* 17 (695), pp. 1–8.

Fuggle, S. (2018). 'Clinical Biochemistry Reference Ranges Handbook'. East Sussex Healthcare (NHS) [online] Accessed 3rd November 2019. URL: `https://www.esht.nhs.uk/wp-content/uploads/2017/08/Clinical-Biochemistry-reference-ranges-handbook.pdf`.

Galama, T. J., Lleras-Muney, A., and Kippersluis, H. van (2018). 'The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence'. *Oxford Research Encyclopedia of Economics and Finance*.

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

Gelman, A. and Imbens, G. (2019). 'Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs'. *Journal of Business and Economic Statistics* 37 (3), pp. 447–456.

Goodman, A., Lamping, D. L., and Ploubidis, G. B. (2010). 'When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children.' *Journal of Abnormal Child Psychology* 38 (8), pp. 1179–1191.

Goodman, R., Meltzer, H., and Bailey, V. (1998). 'The Strengths and Difficulties Questionnaire: A Pilot Study on the Reliability and Validity of the Self-report Version'. *European Child & Adolescent Psychiatry* 7 (3), pp. 125–130.

Greene, W. (2004). 'The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects'. *Econometrics Journal* 7 (1), pp. 98–119.

Grinberg, A. (2015). 'The Effect of Birth Order on Occupational Choice'. *Atlantic Economic Journal* 43, pp. 463–476.

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

Gujarati, D. N. (2004). *Basic Econometrics*. Fourth. McGraw-Hill, pp. 341–386.

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

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

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

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

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

Hanushek, E. A. (1992). ‘The Trade-off between Child Quantity and Quality’. *Journal of Political Economy* 100 (1), pp. 84–117.

Hanushek, E. A. and Wößmann, L. (2006). ‘Does Early Tracking Affect Educational Inequality and Performance? Differences-in-Differences Evidence across Countries’. *Economic Journal* 116 (510), pp. C63–C76.

Harakeh, Z., Engels, R. C. M. E., Vermulst, A. A., De Vries, H., and Scholte, R. H. J. (2007). ‘Psychology and Health The influence of best friends and siblings on adolescent smoking: A longitudinal study’. *Psychology and Health* 22 (3), pp. 269–289.

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

Hausman, J. A. and Wise, D. A. (1979). ‘Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment’. *Econometrica* 47 (2), p. 455.

Heckman, J. J., Pinto, R., and Savelyev, P. a. (2013). ‘Understanding the mechanisms through which an influential early Childhood programm boosted Adult outcomes’. *American Economic Review* 103 (6), pp. 2052–2086.

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’. *Review of Economic Studies* 64 (4), pp. 487–535.

Heckman, J. J., Vytlacil, E. J., and Di, U. (1999). ‘Local instrumental variables and latent variable models for identifying and bounding treatment effects’. *Proceedings of the National Academy of Sciences of the United States of America* 96, pp. 4730–4734.

Ho, D. E., Imai, K., King, G., and Stuart, E. A. (2007). ‘Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference’. *Political Analysis* 15 (3), pp. 199–236.

Hotz, J. V. and Pantano, J. (2015). ‘Strategic parenting, birth order, and school performance’. *Journal of Population Economics* 28 (4), pp. 911–936.

House of Commons (1976). 'Percentage of leavers from schools in England going to university from 1960 to 1974'. Written Answers, House of Commons, Education and Science. [online] Accessed 25th November 2019. URL: `https://api.parliament.uk/historic-hansard/written-answers/1976/mar/29/school-leavers`.

Hyytinen, A., Meriläinen, J., Saarimaa, T., Toivanen, O., and Tukiainen, J. (2018). 'When does regression discontinuity design work? Evidence from random election outcomes'. *Quantitative Economics* 9 (2), pp. 1019–1051.

Iacovou, M. (2008). 'Family Size, Birth Order, and Educational Attainment'. *Marriage & Family Review* 42 (3), pp. 35–57.

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

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

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

Jerrim, J. and Sims, S. (2018). 'The association between attending a grammar school and children's socio-emotional outcomes. New evidence from the Millennium Cohort Study'. *British Journal of Educational Studies (forthcoming).*

— (2019). 'How Do Academically Selective School Systems Affect Pupils' Social-Emotional Competencies? New Evidence From the Millennium Cohort Study'. *American Educational Research Journal* 56 (5), pp. 1769–1799.

Jesson, D. (2000). 'The Comparative Evaluation of GCSE Value-Added Performance in Two-Sector Models of Endogenous Growth by Type of School and LEA'. Economics Discussion Paper No. 52, University of York.

— (2013). 'The Creation, Development and Present State of Grammar Schools in England'. Unpublished manuscript.

Jones, A. M., Pastore, C., and Rice, N. (2018). 'Tracking pupils into adulthood: selective schools and long-term well-being in the 1958 British cohort'. HEDG Working Paper No. 18/32, University of York.

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* 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* 42 (2), pp. 369–394.

Juntunen, K. S., Laara, E. M., and Kauppila, A. J. (1997). 'Grand grand multiparity and birth weight.' *Obstetrics and gynecology* 90 (4), pp. 495–499.

Kaplan, G. A. and Camacho, T. (1983). 'Perceived health and mortality: A nine-year follow-up of the human population laboratory cohort'. *American Journal of Epidemiology* 117 (3), pp. 292–304.

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

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

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

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

Khong, T. Y., Adema, E. D., and Erwich, J. J. H. M. (2003). 'On an anatomical basis for the increase in birth weight in second and subsequent born children.' *Placenta* 24 (4), pp. 348–353.

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

Kumar, S. (2016). 'The effect of birth order on schooling in India'. *Applied Economics Letters* 23 (18), pp. 1325–1328.

Lee, D. S. and Lemieux, T. (2010). 'Regression Discontinuity Designs in Economics'. *Journal of Economic Literature* 48 (2), pp. 281–355.

Lehmann, J.-Y. K., Nuevo-Chiquero, A., and Vidal-Fernandez, M. (2018). 'The Early Origins of Birth Order Differences in Children's Outcomes and Parental Behavior'. *Journal of Human Resources* 53 (1), pp. 123–156.

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.

Levačić, R. and Marsh, A. J. (2007). 'Secondary Modern Schools: Are Their Pupils Disadvantaged?' *British Educational Research Association* 33 (2), pp. 155–178.

Lewis, G., Pelosi, A. J., Araya, R., and Dunn, G. (1992). 'Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers.' *Psychological Medicine* 22 (2), pp. 465–486.

Lynn, P. and Borkowska, M. (2018). 'Some Indicators of Sample Representativeness and Attrition Bias for BHPS and Understanding Society'. Understanding Society Working Paper No. 2018-01, ISER, University of Essex.

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

Maguire, L. K., Niens, U., Mccann, M., and Connolly, P. (2016). 'Emotional development among early school-age children: gender differences in the role of problem behaviours'. *Educational Psychology* 36 (8), pp. 1408–1428.

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

Maurin, E. and McNally, S. (2009). 'The Consequences of Ability Tracking for Future Outcomes'. Unpublished manuscript.

Mayhew, K., Deer, C., and Dua, M. (2004). 'The move to mass higher education in the UK: Many questions and some answers'. *Oxford Review of Education* 30 (1), pp. 65–82.

McCrary, J. (2008). 'Manipulation of the running variable in the regression discontinuity design: A density test'. *Journal of Econometrics* 142 (2), pp. 698–714.

Mechoulan, S. and Wolff, F.-C. (2015). 'Intra-household allocation of family resources and birth order: evidence from France using siblings data'. *Journal of Population Economics* 28 (4), pp. 937–964.

Mincer, J. (1958). 'Investment in human capital and personal income ability'. *Journal of Political Economy* 66 (4), pp. 281–302.

Monfardini, C. and See, S. G. (2016). 'Birth order and child cognitive outcomes: an exploration of the parental time mechanism'. *Education Economics* 24 (5), pp. 481–495.

Moroni, G., Nicoletti, C., and Tominey, E. (2019). 'Child Socio-Emotional Skills: The Role of Parental Inputs'. HCEO Working Paper No. 2019-038, The University of Chicago.

Narusyte, J., Ropponen, A., Alexanderson, K., and Svedberg, P. (2017). 'Internalizing and externalizing problems in childhood and adolescence as predictors of work incapacity in young adulthood'. *Social Psychiatry and Psychiatric Epidemiology* 52 (9), pp. 1159–1168.

Oakes, J. (2005). *Keeping track: How schools structure inequality.* Yale University Press. New Haven and London.

OECD (2016). 'PISA 2015 Results (Volume II) Policies and practices for successful schools'. PISA, OECD Publishing, Paris.

Pavan, R. (2016). 'On the Production of Skills and the Birth-Order Effect'. *Journal of Human Resources* 51 (3), pp. 699–726.

Persson, P. and Rossin-Slater, M. (2018). 'Family Ruptures, Stress, and the Mental Health of the Next Generation'. *American Economic Review* 108 (45), pp. 1214–1252.

Ploubidis, G. B., Sullivan, A., Brown, M., and Goodman, A. (2017). 'Psychological distress in mid-life: Evidence from the 1958 and 1970 British birth cohorts'. *Psychological Medicine* 47 (2), pp. 291–303.

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

Porter, J. and Yu, P. (2015). 'Regression discontinuity designs with unknown discontinuity points: Testing and estimation'. *Journal of Econometrics* 189 (1), pp. 132–147.

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

Price, J. (2008). 'Parent-Child Quality Time: Does Birth Order Matter?' *Journal of Human Resources* 43 (1), pp. 240–265.

Rees, D. I., Lopez, E., Averett, S. L., and Argys, L. M. (2008). 'Birth order and participation in school sports and other extracurricular activities'. *Economics of Education Review* 27 (3), pp. 354–362.

Reid, R., Riccio, C. A., Kessler, R. H., Dupaul, G. J., Power, T. J., Anastopoulos, A. D., Rogers-Adkinson, D., and Noll, M.-B. (2000). 'Gender and Ethnic Differences in ADHD as Assessed by Behavior Ratings'. *Journal of Emotional and Behavioral Disorders* 8 (1), pp. 38–48.

Rodgers, B., Pickles, A., Power, C., Collishaw, S., and Maughan, B. (1999). 'Validity of the Malaise Inventory in general population samples'. *Social Psychiatry and Psychiatric Epidemiology* 34 (6), pp. 333–341.

Rohrer, J. M., Egloff, B., and Schmukle, S. C. (2015). 'Examining the effects of birth order on personality'. *Proceedings of the National Academy of Sciences of the United States of America* 112 (46), pp. 14224–14229.

Rosenzweig, M. R. and Wolpin, K. I. (1988). 'Heterogeneity, Intrafamily Distribution, and Child Health'. *Journal of Human Resources* 23 (4), pp. 437–461.

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

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

Rutter, M., Tizard, J., and Whitmore, K. (1970). *Education, Health and Behaviour.* Longman Publishing Group: London.

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

Schwandt, H. (2018). 'The lasting legacy of seasonal influenza: in-utero exposure and labor market outcomes labour economics'. Discussion Paper No. 12563, Centre for Economic Policy Research.

Shepherd, P. (2013). 'Bristol Social Adjustment Guides at 7 and 11 years'. 1958 National Child Development Study user guide. Institute of Education, University of London.

Skipper, Y. and Douglas, K. M. (2016). 'The impact of a selective entry examination on children's feelings as they approach the transition to secondary school'. *British Educational Research Journal* 42 (6), pp. 945–961.

Staiger, D. and Stock, J. H. (1997). 'Instrumental Variables Regression with Weak Instruments'. *Econometrica* 65 (3), p. 557.

StataCorp (2017). 'Stata Statistical Software: Release 15.' [software] College Station, TX: StataCorp LLC.

Strange, V. (2002). 'What influences peer-led sex education in the classroom? A view from the peer educators'. *Health Education Research* 17 (3), pp. 339–350.

Sullivan, A., Joshi, H., and Leonard, D. (2011). 'Single-sex schooling and labour market outcomes'. *Oxford Review of Education* 37 (3), pp. 311–332.

Sulloway, F. J. (1996). *Born to rebel: Birth order, family dynamics, and creative lives.* New York, NY, US: Pantheon Books.

— (2001). 'Birth Order, Sibling Competition, and Human Behavior'. *Conceptual Challenges in Evolutionary Psychology: Innovative Research Strategies.* Ed. by H. R. Holcomb. Dordrecht: Springer Netherlands. Chap. 2, pp. 39–83.

Tenikué, M. and Tequame, M. (2017). 'Birth Order, Sex Composition and Risky Behaviour of Adolescent Girls in Nigeria'. Working Paper No. 2017-04, LISER.

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* 27 (3), pp. 531–543.

The Guardian (2015). 'Grammar schools and the feeling of failure. Letters.' [online] Accessed 25th November 2019. URL: `https://www.theguardian.com/politics/2015/oct/20/grammar-schools-and-the-feeling-of-failure`.

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

UK Department for Education (2014). 'National curriculum in England: framework for key stages 1 to 4. Statutory guidance.' [online] Accessed 20th November 2019. URL: `https://www.gov.uk/government/publications/national-`

curriculum-in-england-framework-for-key-stages-1-to-4/the-national-curriculum-in-england-framework-for-key-stages-1-to-4.
UK Department for Education (2018). 'Schools That Work For Everyone. Government consultation response.' (September). [online] Accessed 21st November 2019. URL: https://www.gov.uk/government/consultations/schools-that-work-for-everyone.
— (2019a). 'Participation rates in higher education: Academic Years 2006/2007 – 2017/2018'. [online] Accessed 25th November 2019. URL: https://www.gov.uk/government/statistics/participation-rates-in-higher-education-2006-to-2017.
— (2019b). 'Selective Schools Expansion Fund: successful applications, 2018 to 2019. Transparency data'. [online] Accessed 25th November 2019. URL: https://www.gov.uk/government/publications/selective-schools-expansion-fund-successful-applications-2018-to-2019.
University of Essex, Institute for Social and Economic Research (2019). *Understanding Society: Waves 1-8, 2009-2017 and Harmonised BHPS: Waves 1-18, 1991-2009.* [data collection], UK Data Service, 11th Edition, Accessed 11 March 2019. SN: 6614. URL: http://doi.org/10.5255/UKDA-SN-6614-12.
University of London. Institute of Education. Centre for Longitudinal Studies. (2008a). *National Child Development Study: Sweep 4, 1981, and Public Examination Results, 1978.* [data collection]. 2nd Edition. National Children's Bureau, [original data producer(s)]. UK Data Service. SN: 5566. URL: http://doi.org/10.5255/UKDA-SN-5566-1.
— (2008b). *National Child Development Study: Sweep 5, 1991.* [data collection]. 2nd Edition. City University. Social Statistics Research Unit, [original data producer(s)]. UK Data Service. SN: 5567. URL: http://doi.org/10.5255/UKDA-SN-5567-1.
— (2008c). *National Child Development Study: Sweep 6, 1999-2000.* [data collection]. 2nd Edition. Joint Centre for Longitudinal Research, [original data producer(s)]. UK Data Service. SN: 5578. URL: http://doi.org/10.5255/UKDA-SN-5578-1.
— (2012). *National Child Development Study: Sweep 8, 2008-2009.* [data collection]. 3rd Edition. UK Data Service. SN: 6137. URL: http://doi.org/10.5255/UKDA-SN-6137-2.
— (2014). *National Child Development Study: Childhood Data, Sweeps 0-3, 1958-1974.* [data collection]. 3rd Edition. National Birthday Trust Fund, National Children's Bureau, [original data producer(s)]. UK Data Service. SN: 5565. URL: http://doi.org/10.5255/UKDA-SN-5565-2.

Verbeek, M. and Nijman, T. (1992). 'Testing for Selectivity Bias in Panel Data Models'. *International Economic Review* 33 (3), pp. 681–703.

Von Hinke Kessler Scholder, S., Wehby, G. L., Lewis, S., and Zuccolo, L. (2014). 'Alcohol exposure in utero and child academic achievement'. *Economic Journal* 124 (576), pp. 634–667.

Wilcox, M. A., Chang, A. M., and Johnson, I. R. (1996). 'The effects of parity on birthweight using successive pregnancies.' eng. *Acta obstetricia et gynecologica Scandinavica* 75 (5), pp. 453–459.

Wolf, A. (2011). 'Review of Vocational Education: the Wolf Report'. London: Department for Education.

Wooldridge, J. M. (2010). 'Sample Selection, Attrition, and Stratified Sampling'. *Econometric Analysis of Cross Section and Panel Data*. First ed. The MIT Press. Chap. 17, pp. 578–590.

Woolridge, J. M. (2013). 'Advanced Panel Data Methods'. *Introductory Econometrics: A Modern Approach*. Fifth ed. South-Western, Cengage Learning. Chap. 14, pp. 485–501.

World Health Organisation (2019). *Exclusive breastfeeding under 6 months Data by country*. Global Health Observatory data repository. [online] Accessed on 3rd May 2019. URL: `http://apps.who.int/gho/data/view.main.NUT1730?lang=en`.

World Health Organization (2017). *Body mass index - BMI*. [online] Accessed on 24th September 2017. URL: `http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi`.

Young, A. (2019). 'Consistency without Inference: Instrumental Variables in Practical Application'. Unpublished manuscript.