Three Essays on the Economics of Human Capital Development

Emma Louise Gorman

A thesis submitted in partial fulfilment of the requirements for the degree Doctor of Philosophy in Economics

LUMS

Department of Economics

Lancaster University

United Kingdom

December 2018

Three Essays on the Economics of Human Capital Development

Emma Louise Gorman

A thesis submitted to the University of Lancaster in partial fulfilment of the requirements of the degree of Doctor of Philosophy in Economics

December 2018

This thesis is concerned with the role of formal schooling in the production of human capital over the lifecycle.

While many studies have documented an associated between education and cognitive outcomes, less is known about the extent this association is *causal*—or the nature of the underlying mechanisms driving any effect. Chapter 1 examines the causal effect of additional secondary schooling on cognitive function in later life, using new methods in causal mediation analysis to explore the role of occupation choice as a key channel.

The findings reveal robust evidence that basic education leads to improved working memory, but detect little support for effects on verbal fluency or numeric abilities. Staying in school for an additional year increases the probability of entering a higher status occupation, and an analysis of mechanisms finds that up to about one-fifth of schooling's effect on cognitive outcomes can be explained by occupation choice. However, the estimates are too imprecise to yield firm conclusions.

Chapters 2 and 3 are situated in the school choice literature. The promise of school choice is to allow parental preference to influence which school their child attends, weakening the link between residential location and school quality. However, choice is typically constrained—and markets for schools are no exception. Popular schools tend to be oversubscribed, and inevitably many families miss out on a place at their preferred school.

Chapter 2 traces the consequences of missing out on a place at a preferred secondary school in England, focusing on long run outcomes, including high-stakes examination results. The analyses do not find evidence that failing to gain a place at a preferred school leads to poorer academic outcomes—but those who miss out are more likely to engage in risky behaviours, drop out of secondary school, and have poorer mental health in adulthood.

Finally, Chapter 3 assesses the effects of missing out on a place at a preferred primary school, on cognitive and non-cognitive skill development and parental responses. Little evidence is revealed for a detrimental effect on skill development, but compared with those who get into their preferred school, parents whose child misses out on a place are more likely to invest in private tutoring and exam preparation for selective schools.

For my parents

This thesis is submitted in partial fulfilment of the requirements for the degree Doctor of Philosophy in Economics at Lancaster University.

I declare that the thesis has been composed by myself and that the work has not be submitted for any other degree or professional qualification.

I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included. The work presented in Chapter 2 was conducted jointly with Professor Ian Walker: the study design was conceived, planned and revised by both the authors, I carried out the data preparation, data analyses and drafted the manuscript.

Spending the last four years teaching and studying economics has been a privilege, and I am grateful to the Department of Economics at Lancaster and the Economic and Social Research Council for affording me this opportunity. Many people have extended their generous support throughout my PhD candidacy. I couldn't have asked for better supervisors in Ian Walker and Eugenio Zucchelli, who offered their support, good humour and guidance right from day one. Thanks especially to Ian Walker, for his belief in me, my work and my capabilities, his practical and academic support, and continued inspiration. The Department of Economics at Lancaster has provided a collegial and positive environment to be a PhD student: thanks to Caren Wareing, Themis Pavlidis and Orestis Troumpounis for their work supporting the PhD students and the PhD programme. I'm grateful for the friends and colleagues I have met along the way: Ayesha Ali, Christina-Jane Crossman-Barnes, Shane Murphy, Vincent O'Sullivan, Jennifer Robinson, Simon Spavound, Wiktoria Tafesse, Mat Weldon, and many others who have made the journey more colourful. Thanks to my parents, Penny and Peter, for always being there for me—even when I moved to the other side of the world. Finally, thanks to Alfred: academic cheerleader and partner in life.

Contents

Introduction

This thesis studies formal schooling and human capital development over the lifecycle. Since Gary Becker's seminal work in 1964 (Becker, 1964), economists have studied the production of human capital and skill formation. The remit of this area has now broadened to include investment in health capital and socio-emotional skills—in addition to traditional measures of cognitive skill. Recent development have highlighted the importance of a lifecycle perspective: understanding how the efficacy of interventions varies with age, to inform on the optimal timing of policy intervention. This work focuses on primary and secondary school periods—stages of life still eminently open to intervention, with scope for lifetime benefits.

In Chapter 1, I examine the nexus between basic education, occupation choice and cognitive skills. First, I consider the causal effect of an additional year secondary schooling on cognitive function in later life (measured at about 60 years). Second, I use new methods in causal analysis to test the the role of the "use-it-or-lose-it" hypothesis: does working in a cognitively demanding occupation preserve cognitive function? Chapters 2 and 3 turn to focus on school choice: does it matter if children attend the school preferred by their parents? For what outcomes, and when?

We know that cognitive ability is an important determinant of many social, economic and health outcomes (Heckman *et al.* , 2006; Wraw *et al.* , 2015). Given there are large socioeconomic difference in cognitive outcomes, and trajectories of cognitive development, how to narrow this gap and intervene to improve life chances is a common policy focus. Recent studies recognise that the issue is not as simple as "nature vs nurture", as older literature would suggest (Herrnstein & Murray, 1994). Rather, measured skills are a product of a complex interplay between genetics, early-life experiences and skill investments (Heckman, 2006, 1995).

In Chapter 1, I contribute to this area by providing new evidence on whether an additional year spent in secondary school exerts a *causal* impact on cognitive outcomes later in life. I use data from Understanding Society, a large panel dataset following the lives of over 40,000 households in the United Kingdom (UK). I use a Regression Discontinuity Design to exploit a change in school leaving laws, building on a growing literature using policy experiments to credibly inform on the wider returns to schooling (Glymour *et al.* , 2008; Banks & Mazzonna, 2012). In line with the findings of other studies, the main results show that an additional year of schooling leads to improved working memory in later life, with large effect sizes ranging from about one third to half of a standard deviation. However, the findings reveal little evidence for causal impacts on verbal fluency or basic numeracy skills.

A second key question is then about *how* extra schooling causally could improve cognitive outcomes, and the second contribution of Chapter 1 is to examine the plausibility of a specific potential channel: occupation choice. Occupation is an commonly studied factor in preserving our cognitive health—creating resilience to the wear-and-tear on our brain associated with aging (termed cognitive reserve in the neuroscience literature (Whalley *et al.* , 2004)). This motivates the oft-cited "use-it-or-lose-it" hypothesis: continuing to engage in cognitively stimulated work, or work-like activities, is one way to protect our cognitive abilities in older ages (Rohwedder & Willis, 2010).

An intervention can exert causal effects through many different mechanisms, and increasingly, economists are recognising the important of understanding not just *whether* an intervention works, but also how. Understanding how an intervention works is key for assessing external validity: whether the policy would be effective in a different setting. The importance of studying mechanisms is highlighted by Angus Deaton in a discussion the utility of an average treatment effect, denoted θ (Deaton, 2010, p.p. 429):

> "*[...] there are many possible mechanisms, some of which will work in one context and not in another. In consequence θ is unlikely to be constant [...] In-*

> *stead, it is precisely the variation in θ that encapsulates the poverty reduction mechanisms that ought to be the main objects of our enquiry.*

One reason the study of mechanisms is relatively scarce is because it is difficult to do credibly. Studying the role of a post-treatment variable introduces a second endogeneity problem, in addition to endogeneity of the treatment. A full structural model is one way to study mechanisms, however this is not without its own stringent data and assumption requirements. A more common method is to take a data-driven approach and employ an effect decomposition. Yet, with some exceptions, *identification* of mechanisms explaining treatment effects has not received much attention in the literature, as highlighted by Fortin *et al.* (2011, p.p. 13) in their review of decomposition methods in economics:

> *"In econometrics, the standard approach is to first discuss identification [...] and then introduce estimation procedures to recover the object we want to identify. In the decomposition literature, most papers jump directly to the estimation issues (i.e., discuss procedures) without first addressing the identification problem."*

I confront this concern directly by applying newly developed methods in the causal mediation analysis literature. This is an area of study which has developed recently, in the econometrics and statistics literature, to formally disentangle causal mechanisms driving treatment effects. An important point of difference in this literature is to make explicit the identification assumptions required to decompose average treatment effects into their constituent *causal* channels, and indeed to highlight the inherent difficulty of this task (see Huber *et al.* (2016), Keele *et al.* (2015), Imai *et al.* (2011) and Huber (2016) for useful overviews of identification and estimation of direct and indirect treatment effects).

Applied examples of these methods in the economics literature include: Huber (2015), who uses inverse probability weighting to disentangle the role of occupation choice in explaining wage gaps; Huber *et al.* (2017a), who deploy instrumental variables to study the causal mechanisms driving the education-income gradient; Huber *et al.* (2017b), who consider the role of tougher caseworkers as causal mechanism connecting active labour market programs and increased employment; and Deuchert *et al.* (2018), who quantify causal channels shaping the effects of the Vietnam draft in a difference-in-difference setting.

In the Understanding Society data, I observe the occupation chosen both immediately after leaving school, as well in toward the end of working life. I use two classifications of occupation type. The first classification splits occupations up from the most routine tasks up to professional and managerial tasks, which tend to offer both more autonomy as well as safer working environments. Second, I classify the occupations based on how cognitively-intensive they are—the extent to which they rely on "STEM"-type skills, technical skill, critical analysis and so on.

I apply the methods for causal mediation analysis developed in Yamamoto (2014); Keele *et al.* (2015); Park & Kürüm (2018), which decompose a Local Average Treatment Effect (LATE) into a direct and indirect effect. This approach exploits the raising of the school leaving age as an instrument for endogenous schooling. In the absence of a credible instrument, I rely on a type of conditional independence assumption to handle the endogeneity of occupation choice. This is paired with a sensitivity analysis to assess the robustness to deviations from this assumption.

The analyses of mechanisms shows that staying on for an extra year of school increases the probability of entering a higher status job after leaving school. Second, it shows that occupation type could explain up to 18% of the total effect of schooling on memory (depending on which measure is used), a similar figure to that found in other studies. However, the figures are very imprecisely estimated, and not statistically significant at conventional levels. Given the imprecision of the estimates, I remain agnostic about the role of occupation, but conclude that basic education causally improves working memory—an important component of cognitive function in older ages.

Evidently, quantity of schooling matters for some important outcomes. But what about school quality? Chapters 2 and 3 assess the effects of *which* school a child attends. Specifically, Chapters 2 and 3 trace the consequences of missing out on a place at the school which is most preferred by parents in England. These chapters contribute to the debate around "school choice". School choice policies are policies which allow parental preferences to influence which school their child will attend (Cantillon, 2017). Despite the innocuous-sounding definition, school choice has been hotly debated and politicised issue both in England and internationally. In England, features of parental choice and a "quasi-market" in compulsory schooling were introduced largely from 1988, in the 1988 School Reform Act. Indeed many education systems around the world feature some aspect of parental choice, and today the policy discussion is as much about *how* to organise school choice and admissions, as about whether to have parental choice at all.

The central problem in designing a school choice system is allocating pupils to schools, accounting for parental preferences, school capacities and wider policy objectives (e.g. diversity in schools, or social mobility). One motivation for school choice focuses on recognising heterogeneity of parental preferences, by allowing families to express preferences over a diversity of school types. A second, and perhaps more commonly cited, motivation, and not necessarily implied by the first, is to induce competition between schools, with the aim of improving academic standards. Both of these rationales have been referenced by different UK Governments in relation to school choice policy.

Understanding how parents value schools, and the effects of attending a preferred school, are important for at least two reasons. First, understanding what parents value in schools is important for meeting to stated aims of a school choice policy. For example, if governments value academic performance in schools, but parents value something else (e.g., proximity to home, anti-bullying policies), a policy of parental choice will not have the desired effect of raising standards via increased competition on academic performance. Second, if attending a preferred school matters for life chances, then from an equity perspective it matters *who* gets in. Presently, there is substantial variation in rates of preferred school attendance by region, by ethnicity and by socio-economic status—understanding why this variation exists and the possibly heterogeneous consequences is important for developing policy.

The promise of school choice programs is to break the connection between where families live and the quality of the school they attend. The limitation of the traditional approach—allocating children to their local school—amplifies the social stratification associated with Tiebout sorting: that is, affluent neighborhoods tend to enjoy higher quality public amenities (Tiebout, 1956). However choice is seldom unconstrained, and schools are no exception. Schools perceived to be good tend to be oversubscribed, such that not all pupils can attend their preferred school. Each year, about 16% of families miss out on a place at their preferred secondary schools, and about 6% miss out on a place at their preferred primary school. Whilst attracting much ire from the media and dissatisfied parents, there is little evidence about the consequences of missing out. Hence, Chapters 2 and 3 in this thesis are specifically concerned with *effects of missing out on a place at a preferred school*.

Internationally, various incarnations of school choice policies are common, and a growing literature aims to characterise the nature of parental preferences over schools and the consequences of gaining a place at a school preferred by parents (Pop-Eleches & Urquiola, 2013; Cullen *et al.* , 2006; Dobbie & Fryer, 2014, 2015; Abdulkadiroglu *et al.* , 2017b, 2014). The research on preferences report much heterogeneity. In the USA, recent research has shown that parents often value *peer mix* (i.e., wealthy white peers) rather than school effectiveness (i.e., a causal improvement in test scores) (Abdulkadiroglu *et al.* , 2017a).

Indeed, comparisons of the outcomes of pupils who just get into a prestigious school, with those who just miss out reveal no impact on short run test scores, despite large discontinuity in peer mix and school prestige (Abdulkadiroglu *et al.* , 2014). With some exceptions, this finding is common across many settings: gaining a place a parents' preferred school does not appear to causally influence short run test scores Abdulkadiroglu *et al.* (2014); Dobbie & Fryer (2014). There is some evidence of improvements in broader outcomes, such as risky behaviours and truancy (Dudovitz *et al.* , 2018), later employment rates, and timing of childbearing (Jackson & Beuermann, 2018).

In Chapter 2, a jointly authored study with Ian Walker, we use data from the Longitudinal Study of Young People in England, (LSYPE), confidentially linked to high quality administrative data on school attributes and educational outcomes (National Pupil Database) to assess the effects of missing out on a preferred secondary school on a range of high-stakes outcomes. The central challenge is to compare families who had similar chances of admission to a school and similar preferences over schools, yet one gains a place and one misses out. For those who miss out on their preferred school, we do not have data on the counterfactual school they would have preferred to attend, and this must be estimated from the data.

We employ a selection-on-observed variables strategy, informed by the nature of the institutional setting (to account for admissions probability) and literature on parental preferences (to account for preferences) and information on the local population and school type (to account for variation in choice sets). We use a combination of matching and regression to compare outcomes of children with similar values of the variables shaping admissions probability, similar variables which shape preferences, a similar feasible choice set of schools, and similar prior ability and socio-economic status.

While preferred schools have higher headline measures of attainment (e.g., GCSE results), our analysis fails to detect convincing evidence for effects of missing out a preferred school on test scores and university attendance. However, we do find consistent evidence of increased high school drop-out after the age of compulsory schooling, and increased university drop-out among pupils who miss out on a preferred school. These effects persist to age 25 years, where we find long-lasting negative effects on mental health and income.

Pupils who miss out show increased patterns of early engagement in risky behaviours, representing a potential mechanism driving the increased drop-out rates. While we do not find strong evidence for gender differences, the negative consequences of missing out are more pronounced in local areas which used a more manipulable mechanism ('first-preference first') for allocating school places. Pupils exposed to the first-preference mechanism who miss out end up commuting further to school than their counterparts exposed to a Deferred Acceptance mechanism, possibly indicating much poorer school matches. Overall, our findings highlight that which school a child attends can have consequences for broader outcomes than only test scores, and the way places are allocated matters for ensuring children attend a school which is right for them.

Chapter 3 moves to look at the consequences of not attending a preferred *primary* school, tracing pupil's cognitive and non-cognitive skill development and the dynamics of parental behavioural responses. I employ data from the Millennium Cohort Study (MCS), a cohort study following the lives of around 18,000 children born in during the years 2000 and 2001. Primary school admissions are governed by similar admissions rules to secondary schools, and I use an analogous identification strategy to that used in Chapter 2—i.e., a selection-on-observed variables strategy implemented using propensity score matching.

In contrast to the LSYPE, the MCS directly elicits parents' preferences over schools. Specifically, it asks about the identity counterfactual school by asking: if your current school was not your first choice school, which school was your first choice? From this question, we can compute the distance from a family's home address to their current school, and crucially, the distance to their *preferred school* (if they were not offered a place at their first-choice school). As a matching variable, distance to preferred school is especially informative because it capture both admission probability, as well as preferences—a family's willingness-to-travel for a preferred school.

The results do not reveal any effects of missing out on a first choice school on cognitive development. In terms of parental responses, there are small (but statistically insignificant) increases in the probability of parental investment activities at ages 7—including helping with homework and paying for private tuition. At age 11 years, parents of children who miss out are significantly more likely to pay for extra lessons. This work does not detect any evidence that attending a non-preferred parental primary school has any lasting effects on the skill development of children. However, the findings show that parental responses are important when considering policy impacts.

Taken together, this thesis reports on new research assessing the role of schools in promoting the development of human capital. Using modern methods of policy evaluation, I explore the consequences of two key aspects of education: (1) how many years a child spends in secondary school, and (2) which school they attend, and whether this is they school preferred by parents. Increasingly, education and skills are the key to getting on in life, in terms of both labour market success and wider life outcomes. In the face of a fundamentally changing economy, understanding the role of the school system in promoting human capital development remains an essential area of study for effective policy design. This thesis claims the modest aim of making a useful contribution to this agenda.

Chapter 1

The causal effects of secondary schooling on cognitive function

This work was funded by the an ESRC PhD studentship "The long-term effect of education on health: New evidence from the United Kingdom". I am grateful for helpful discussions and comments from my supervisors Ian Walker and Eugenio Zucchelli, and participants at: the North West Doctoral Training Economics PhD Conference 2016; American Society for Health Economists Biennial Conference 2016; IZA Summer School in Labor Economics 2017; NWSSDTP PhD Conference in Economics 2017; International Health Economics Association World Congress 2017. Thank you to Andy Dickerson for providing the O*NET data.

1.1 Introduction

Cognitive performance shapes economic, social and health outcomes over the lifecycle (Heckman *et al.* , 2006; Wraw *et al.* , 2015). While the efficacy of early intervention in improving cognitive performance is well studied, less is known about whether these outcomes are malleable outside of early childhood (Heckman, 2006). This study assesses whether secondary schooling has lasting causal effects on cognitive function—measured four decades after leaving school—and explores the role of occupation type in shaping these effects.

In later life, cognitive performance becomes an important component of healthy ageing and continued independent functioning (Beard *et al.* , 2015). Among healthy individuals—free of clinical cognitive impairment—age-related reductions in cognitive performance can impede the ability to carry out daily activities (Tucker-Drob, 2011; Boyle *et al.* , 2012) and manage financial planning decisions (Hsu & Willis, 2013). This has important implications for continued labor market engagement and retirement planning, as state pension ages rise and supporting continued labour market participation is a central policy focus (OECD, 2006; Department for Work and Pensions, 2013).

Declining cognitive performance is also a risk factor for morbidity and mortality. One example is dementia, for which declines in cognitive function, especially memory, are a precursor. Estimates from 2016 showed that 800,000 people suffered from dementia in the United Kingdom, with a total cost of 26 billion pounds per year—largely comprising informal care costs (Prince *et al.* , 2014). Even after accounting for the observed and projected reductions in age-specific incidence rates, the prevalence of dementia is forecast to increase to 1.2 million by 2040, representing a vast increase in both human and economic costs (Ahmadi-Abhari *et al.* , 2017).

Yet cognitive impairment is not an inevitable consequences of ageing, and a large body of work has sought to uncover modifiable risk factors across the lifecourse. Basic education is one promising candidate in improving cognitive outcomes. Based on observational evidence, *The Lancet*'s recent dementia Commission concluded that 8% of dementia cases could be avoided by increased levels of basic education (Livingston *et al.* , 2017). In the same vein, age-specific rates of incident dementia are falling (Ahmadi-Abhari *et al.* , 2017), predominantly among those with at least a high school diploma (Chêne *et al.* , 2016).

Although causal evidence is sparse, these findings are consistent with the hypothesis that increased population levels of education could reduce the growth in burden of cognition-related disease, and support economic adjustment to a changing demographic structure. With this in mind, the objectives of this study are two-fold. First, to provide new evidence on the extent to which schooling shapes cognitive function in later life. Second, to conduct a formal analysis exploring the channels driving these effects, paying particular attention to the role of occupation type and complexity.

Despite extensive observational research demonstrating that schooling is a correlate of cognitive outcomes throughout the lifecycle (Plassman *et al.* , 1995), fewer studies are able to attach a causal to this association or the hypothesised underlying mechanisms. One approach, leveraged in this paper, is to exploit the substantial investments in publicly provided education occurring over the last century, notably the successive increases in the minimum age at which students can leave secondary school. These increases in the minimum schooling leaving age represent plausibly exogenous changes in schooling, and are widely employed to quantify both pecuniary and non-pecuniary returns to education (Harmon & Walker, 1995; Oreopoulos & Salvanes, 2011). The cohorts affected by these reforms are now ageing, providing an opportunity to explore whether the effects of schooling persist into later life.

The United Kingdom (UK) provides an especially informative laboratory to examine the causal effects of schooling. Two changes to the secondary school leaving age were enacted during the 20th century: in 1947 the minimum school leaving age was raised from 14 years to 15 years, and 1972 it was raised again to 16 years. These changes affected relatively large shares of the relevant cohorts. For example, the corresponding changes in the United States exerted a causal effect on only about five percent of the student cohort, compared with about one half and one third for the 1947 and 1972 reforms in the UK, respectively. Yet, the evidence base on schooling and cognitive function in the United Kingdom remains limited, and little is known about the underlying mechanisms driving these long-run effects.

For England and Wales, Banks & Mazzonna (2012) established positive effects of an additional year of schooling on aspects of cognitive function among the cohorts affected by the 1947 raising of the school leaving age. The findings are large in magnitude, raising the possibility that increased population education could function as a policy tool in fostering healthy ageing. Given the large total effect size, the protective effect of education of cognition is unlikely to be wholly a direct effect. More plausibly, education influences cognition indirectly, through the wider set of opportunities that education affords. However, the cohorts exposed to the 1947 reform grew up in unstable economic and political circumstances; their choice set and constraints they faced were vastly different from more recent cohorts. Whether the channels through which that reform operated are still relevant in the current policy context, despite the changing nature of the work tasks, labour market participation, and health technology, remains unclear.

The first contribution of this study is to provide new evidence on the causal effect basic education on cognitive outcomes, using a Fuzzy Regression Discontinuity design to exploit the raising of the secondary school leaving age in 1972. This reform increased the minimum school leaving age from 15 to 16 years in England and Wales. The data employed is from Understanding Society (U.S.)—the largest household panel study in the United Kingdom. This data contains granular information on month-year of birth, allowing exact identification of exposure to the reform. Comparing the outcomes of observationally similar individuals exposed and unexposed to the reform—born only months apart—yields credible estimates of the causal effects of the additional year of schooling.

One threat to the validity of this approach is confounding from secular trends in longevity and cognition: cohort-specific cognitive performance has been steadily increasing over the 20th-century (Flynn, 1987). This so-called *Flynn effect* may be influenced not only by investments in mass education during this time, but the changing nature of job-tasks to be more cognitively stimulating, improvements in nutrition, and the increasing proportion of women entering the labour force, which act as confounding factors (Skirbekk *et al.* , 2013; Case & Paxson, 2009). To reduce concerns of this type of confounding, this study adjusts for birth cohort trends, and uses a small sample window—of just under three years either side of the reform date—to enhance the comparability of the treatment and control units. Additionally, sensitivity analyses are conducted using different covariates sets and sample window choices.

The second contribution is to explore the role of occupation choice as a key mechanism driving the effect of education on cognition. Despite many theories, less is known empirically about the channels which drive the education-cognition gradient. A number of hypotheses highlight the role of occupation type, especially with regard to occupational complexity, and the commonly cited "use-it-or-lose-it" hypothesis. This framework suggests mental stimulation may sustain brain function, supporting the idea that more cognitively stimulating activities stave off cognitive decline (Rohwedder & Willis, 2010). In a similar vein, proponents of the so-called "cognitive reserve" theories in neuropsychology have put forward education and occupation as key factors in increasing mental resilience, reducing the clinical manifestations of brain ageing (Stern, 2002).

This paper applies methods from the causal mediation analysis literature to explore the role of occupation type as a key mechanism (Keele *et al.* , 2015). To allow for potential endogeneity of occupation type—despite the absence of a credible second instrument—the approach used relies on a type of conditional independence assumption for occupation choice (Yamamoto, 2014; Park & Kürüm, 2018). While fundamentally untestable, this assumption is arguably plausible in this application. Sensitivity analyses are conducted to provide support for this assumption, by relaxing the conditional independence assumption and assessing how this impacts the findings, and testing the stability of the effects to the addition of further covariates.

The key findings include a local average treatment effect (LATE) of one additional year of schooling on working memory, ranging from one- to two-thirds of a standard deviation, depending on the sample and model specification employed. This effect is significant at at least the 10% level across a range of reasonable specifications, and is in excess of the Ordinary Least Squares (OLS) estimate of 0.15 standard deviations. While the OLS estimates shows a positive association between schooling and numeric ability and verbal fluency, there is negligible evidence for a causal effect on these outcomes.

Second, the effect size from the analyses of mechanisms indicate occupation type could explain up to 18% of the total effect of schooling on memory, depending on the measure of occupation used, and the model specification. However, the figures are very imprecisely estimated, and not statistically significant at conventional levels. Given the imprecision of the estimates, I remain agnostic about the role of occupation, but conclude that basic education causally improves working memory—an important component of cognitive function in older ages. Increasing population levels of education may be an effective tool in reducing growth in burden of cognition-related disease, and supporting economic adjustment to an ageing population.

1.2 Related literature

A recent empirical literature has used changes in compulsory schooling laws (CSLs) to examine the lasting health and mortality effects of education, including mortality, self-assessed health, disability and health behaviours (recent reviews include Lochner (2011) and Mazumder (2012)). Fewer studies in this area have focused on cognitive outcomes. The first paper to exploit changes in the minimum school leaving age to explore cognition and mental outcomes was Glymour *et al.* (2008), who used state-level variation in the United States to obtain instrumental variable estimates of the effect of schooling on working memory and mental status among cohorts born between 1900 and 1947. These changes in mandated schooling yielded large improvements in performance on memory tests in old age, although no effects were detected for mental status.

Evidence from England and Wales exploited the 1947 school reform, which increased the minimum school leaving age from 14 to 15, Banks & Mazzonna (2012) found a positive effect of an extra year of schooling on the memory and executive function of older men. A number of studies have exploited both time-series and geographical variation in school leaving ages across Europe, using the Survey of Health, Ageing and Retirement in Europe (SHARE). Schneeweis *et al.* (2014) found a positive effect of an extra year of education on memory, and used the panel aspect to reveal evidence of a protective effect against declines in verbal fluency with age. In addition to increased memory scores among older men, Mazzonna (2014) also found reduced probabilities of depression and improved self-reported health. Analyses of potential mechanisms suggested an important role for working conditions in explaining these effects. These patterns are not limited to Europe and the U.S.: Huang *et al.* (2013) leveraged educational differences in primary school completion due to China's Great Famine, during 1959-1961, to assess education effects on cognitive outcomes in older ages. The results revealed a protective effect of cognition, especially episodic memory, of up to 20%.

Generally, the results of these studies find Local Average Treatment Effects estimates in excess of the OLS estimates. Crespo *et al.* (2014) use the SHARElife data—the third wave of the SHARE panel study which contains retrospective life history data—to shed light on the early-life characteristics of those individuals whose behaviour was altered by the implementations of the CSLs, and who may have different returns to schooling than the wider population. The authors find larger effects of the reforms on years of education for those with lower socio-economic status (measured by reports of living in a dwelling with two or fewer rooms), and those reporting better childhood health status. Overall, their results show large, positive effects of extra schooling on memory and depression. Stratifiying the sample by early-life characteristics revealed some variation in point estimates of the causal effect of education, although these differences were not statistically significant.

1.2.1 Schooling and cognition: Potential channels

Given the large total effect size found in many studies, the protective effect of education of cognition is unlikely to be wholly a direct effect. More plausibly, education influences cognition indirectly, through the wider set of opportunities that education affords. The set of plausible channels through which protective effects of schooling on later-life functioning could operate. A recent review of empirical studies identified self-reported health, biomarkers of physiological health (e.g., inflammation), cardiovascular disease and its risk factors, education and occupational trajectories, among others, as risk factors for age-associated cognitive decline (Deary *et al.* , 2009). Certainly a number of channels are likely to operate, for instance income effects, health behaviours and social connections. The empirical evidence for a causal effect of education on health outcomes and behaviours is mixed; many, but not all, credible studies fail to detect any effect (see Clark & Royer (2013); Lochner (2011)). Although the present study focuses on occupation and education, many other factors also play an important role.

This paper aims to test the role of occupation type and labour market engagement. The “cognitive reserve" is a key framework from the neurological literature (Stern, 2002), relating education, occupation and cognitive outcomes. Cognitive reserve is a hypothetical construct used to explain how, facing similar neurodegenerative changes, substantial variation is observed in cognitive ageing, and diagnosed cognitive impairment, across individuals. In other words, individuals do not experience the same declines in everyday cognitive functioning for a given decline in brain health. Cognitive reserve is the concept which is proposed to explain this resilience. Education and occupational complexity are often cited as important proxies for cognitive reserve, in addition to other “lifestyle factors” such as dietary habits and physical exercise. The hypothesis suggests that, for individual with a greater level of cognitive reserve, a greater level of pathology may be required to result in clinical manifestations of any diagnosable impairment.

Second, in the economics literature, a relevant framework is the dynamic model of the demand for good health proposed in Grossman (1972). This model describes the life-cycle accumulation and decumulation of a health stock, and how this relates to choices of occupation, activity and consumption. The implications of this framework can be extended more generally to a stock of cognitive capital: cognitive function in older ages is a component of health, which can be augmented by health investment and cognitive repair inputs, and depreciates with age. The role for education in the seminal Grossman model is to increase the efficiency of health production: those with more education enjoy a higher marginal productivity of a given health input. For example, correct adherance to a medical regimen among more educated individuals. However, evidence for this “productive efficiency" channel in explaining health inequalities is sparse.

Extending this framework, Muurinen (1982) and then Case & Deaton (2005) proposed an augmented model, in which education indirectly reduces the rate of depreciation of the health stock via (a heavily constrained) choice of occupation, activity and consumption. This explanation is more consistent with the wider epidemiological literature on the fundamental causes health inequalities, compared with the "reduced health productivity" argument (Muurinen & Grand, 1985).

In this model, individuals are endowed at birth with stocks of health capital, human capital and financial capital, which are partially substitutable in producing earnings and utility. Those with less human capital will be restricted—via a constrained optimal choice—to occupations which rely more on manual effort or repetitive routine tasks, rather than cognitive skills, to produce earnings. The consequence of this choice is a increased rate of decline in functioning across the lifecycle. Second, a similar effect can be derived through consumption choices. Rather than trading health for earnings through the labour market, one can augment the utility function such that individuals gain utility directly from unhealthy behaviours, again the price for which is paid for directly through a higher rate of health depreciation.

One implication of the model is that the correlation between cognition and education is partly spurious. This framework predicts a higher health stock for a lower rate of time preference and higher initial stock of cognitive or financial assets. This is consistent with empirical regularity that at least 50% of variation in cognitive ability in old age (measured at about 80 years) is explained by cognitive ability in childhood Deary *et al.* (2009). This implication highlights the importance of addressing potential endogenity of education choices.

Focussing specifically on cognitive capital, Mazzonna & Peracchi (2012, 2014) extend this framework to analyse the age-profile of cognitive performance and how this changes after retirement. Empirical tests of the theory show cognitive function generally declines after retirement, as the monetary incentive to invest in this stock is reduced. However this effect importantly depends on occupation type: those employed in physically demanding jobs experience a short-run positive effect of retirement on cognitive function. Moreover, education is found to shape heterogeneity in both the level and age-related decline of cognitive function. Together these findings highlight the potential for occupation type as a mechanism in explaining variation in later-life cognitive function.

The characteristics of higher status occupations entered into through education may not only reduce physical wear and tear, but may also aid in sustaining mental outcomes and cognitive performance—for example, work tasks may be more cognitively demanding and have greater autonomy. These frameworks have motivated a large empirical literature in neuro-epidemiology, exploring associations between many lifestyle factors and cognitive ageing (Deary *et al.* , 2009). One robust predictor of age-related cognitive decline is cardiovascular health. Current smoking status and physical activity are also commonly identified as important predictors—likely through their effects on cardiovascular health. Smoking may be especially promising, as it has also been identified in the economics literature as as an outcome which is causally affected by education (de Walque, 2007; Grimard & Parent, 2007), and as a potential mechanism underlying the education-health gradient (Brunello *et al.* , 2015).

1.3 Data and setting

1.3.1 School reforms in England and Wales

Many studies have leveraged the 1972 raising of the school leave to study the returns to education (e.g., Oreopoulos (2008); Clark & Royer (2013); Dickson *et al.* (2016)). The 1944 Education Act increased the minimum age pupils can leave secondary school from 14 to 15 years, enacted from April 1st 1947; this Act also conferred powers to raise the minimum school leaving age again to 16 years. In March 1972, it was decided that the minimum school leaving age would be raised to 16 years for school cohorts beginning on the 1st of September 1972 (Woodin *et al.* , 2012). This change to 16 years affected birth cohorts of pupils born from the 1st of September, 1957, onward. The 1972 change, aside from being more recent, was enacted in less unstable times—the cohorts exposed to the 1947 change were were affected by the aftermath of the Great Depression, and effects of the World War. The second, 1972, school reform is exploited in this paper. Considering this more recent reform may shed light on whether the results from previous studies using the 1947 change are generalisable, given the particularities of the context faced by those earlier cohorts.

In addition to an extra year of school completed, the 1972 change led to greater rates of completion of formal qualifications; many more students stayed until the end of their 16th school year to obtain “O-level" qualifications. This increase in the probability of gaining a qualification is important. Dickson & Smith (2011) exploit the Easter Leaving Rule—an institutional rule providing exogenous variation in qualification attainment—confirming that some of the observed wages returns to an extra year of schooling in 1972 were due to the increase in qualifications, rather than solely the length of time in school.[1]

1.3.2 Data sources

The dataset employed is Understanding Society (US), a panel study of households in the United Kingdom (McFall, 2013b; Institute for Social and Economic Research & NatCen Social Research, 2015), the successor study to the British Household Panel Survey (BHPS). It commenced in 2009 as a representative probability sample of approximately 40,000 households. Participants are interviewed annually, but each wave of data collection spans 24 months (therefore the survey waves are overlapping and comprise approximately two years of data collection). The “total sample” comprises multiple subsample components: the main General Population Sample (GPS), continuing BHPS members, and the Ethnic Minority Boost (EMB) subsample. Wave 3, beginning on 7th January 2011 and ending on the 12th July 2013, included a Cognitive Ability module which measured self-rated memory, performance on tests of word recall, numeric ability and verbal fluency measures (McFall, 2013a). Wave 3 has a cross-sectional response rate of 61.3%. Throughout the analyses the survey weights provided with the data were employed, to allow for the possibility of endogenous sampling design and response probabilities. The weights ad-

[1] In this work, I have also endeavoured to exploit the Easter Leaving Rule as an instrument to disentangle the effect of an additional year of schooling into a sheepskin effect and a human capital effect. However, in my data the first stage was too weak, resulting in inflated IV estimates.

just for the complex survey design, and combined probabilities of being selected into the BHPS, GPS and EMB and continuing to Wave 3 of the survey.

1.3.3 Variable construction

Years of schooling and cognitive outcomes

The measure of schooling employed is the report of years of secondary schooling completed. This is derived from the schooling question asking "How old were you when you left school?". Then years of schooling is constructed as the age the respondent reports leaving school minus five.

Episodic memory—retrieving memories associated with specific events—was measured by performance on an immediate and delayed word recall test. Memory is an important early indicator of potential cognitive impairment, and word recall tests are used in cognitive impairment screening tests (Kim *et al.* , 2014). A list of 10 words are read aloud by the computer, the respondent is to repeat these words in any order and the number correctly recalled is recorded by the interviewer. This procedure is repeated (about 5 minutes) later in the module to measure delayed recall. Immediate and delayed word recall scores are highly correlated with clinical dementia diagnoses (Wu *et al.* , 2013). The number of words correctly recalled in each test were summed, to create a variable (Word Recall) ranging from 0 to 20.

The Serial 7 Subtraction test is a component of clinical screening instruments for cognitive impairment (i.e,., the Mini Mental State Examination (Crum *et al.* , 1993) and the Cambridge Mental Disorders of the Elderly Examination (Roth *et al.* , 1986)). In the Serial 7 test, respondents are asked to begin at 100 and subtract 7, five times. After each subtraction, the respondent is again prompted by the interview to "take 7 away from that?". The number of correct answers was summed to create a variable ranging from 0 to 5.

The Verbal Fluency test asked respondents to name as many animals as possible in one minute. In addition to testing semantic memory, this tests also executive function as, to perform well, it requires some level of abstract thinking and mental flexibility within a time limit. The test is from the cognitive assessment component of the Cambridge Mental

Disorders of the Elderly Examination, an interview procedure for the diagnosis and measurement of dementia in the elderly (Roth *et al.* , 1986), and has been successfully been employed in extensions of the MMSE (Kim *et al.* , 2014). The number of animals listed ranges from 0 to 71.

The Numeric Ability test assessed ability to solve "every day" numerical problems. The Numeric Ability test asks the respondent five questions of increasing complexity, for example, the first question is as follows: "In a sale, a shop is selling all items at half price. Before the sale, a sofa costs £300. How much will it cost in the sale?". The number of questions correctly answered was summed to have a range of 0 to 5.

The distributions of the raw scores are presented in Appendix 2.A. The continuous measures, Word Recall and Verbal Fluency, were standardised by subtracting the sample mean and dividing by the sample standard deviation, to faciliate interpretation and comparability with previous studies. The Serial 7 Subtraction and the Numeric Ability tests were dichotomised to create a binary variables. *Subtraction* takes the value 1 for respondents with 5 correct answers to the Serial 7 Subtraction, and zero otherwise. *Numeracy* takes the value 1 for respondents with 4 or 5 correct answers to the Numeric Ability test, and zero otherwise.

Occupation variables

The mechanisms underlying the education gradient in cognitive function are inevitably complex and interacting. This paper aims to test one specific hypothesis suggested by theory, namely the role of occupation type and labour market participation patterns. in explaining the effect of schooling on cognitive performance, rather than aiming to exhaustively model all the channels which may be involved.

The first measure employed is the five-class National Statistics Socio-economic Classification (NS-SEC) occupational classification. The five-class NS-SEC groups occupations defined by the Standard Occupation Classification into five categories: 1) Semi-routine; routine; never-worked and long-term unemployed; 2) Lower supervisory and technical occupations; 3) Small employers and own account workers; 4) Intermediate occupations; 5) Higher professional; large employers, higher managerial and administrative occupations.

The NS-SEC is available for both the first job chosen after leaving secondary school, and the current job (or previous job, if the respondent is currently out of the work). The chief focus is on the current, or most recent, occupation, which will capture effects associated with the so-called "use-it-or-lose-it" hypothesis: continuing to engage in a stimulating occupation maintains cognitive reserve and stave off cognitive decline. As an additional measure, the first job is of interest as mechanism because the activity chosen immediately after secondary school is especially amenable to policy intervention. For instance, in the UK context, a current policy development has increased the minimum "participation age", for which young people must remain in education or training, from 16 years to 18 years. This initial start may have longer run implications for labour market trajectories.

Previous occupation was disproportionately missing among those reporting they were long-term sick or disabled, or unemployed. For this group, information from the US employment history module was used to ascertain their activity. Respondents to the US report an employment history of their economic activity status (e.g., full-time work, part-time, unemployed, receiving government benefits, etc.) from when they first left full-time education until their current spell. For those who did not report a current or previous labour market status, individuals who also reported being out of the paid work for at least 50% of their potential working life were coded as NS-SEC category 1 (Semi-routine; routine; never-worked and long-term unemployed).

In addition to the occupation type of first and current job, a final measure of occupation type relates directly to the usage of specific skill types, employing occupational skills profiles developed by Dickerson *et al.* (2012). Dickerson *et al.* (2012) have matched the UK SOC 2010 codes with the US Occupational Information Network (O*NET), a database characterising the skills, abilities, work characteristics of occupations. The O*NET questionnaire is completed by workers and external job analysts. Over 250 questions are asked about job characteristics, or *descriptors*, arranged into the following domains: education and training, knowledge, skills, abilities, work activities, work context and work style. For each descriptor, both the level and importance of that item are elicited. The individual responses are averaged within occupations. Therefore, for each individual, they do not necessarily reflect the skills actually used in a job.

The literature on occupation, education and cognitive outcomes has raised the hypothesis that especially demanding or complex mental tasks may play an important role in maintaining cognitive performance (Fisher *et al.* , 2014). One specific example of these types of complex skills (chosen partly due to data availability) are those used in what are commonly characterised as STEM occupations. These roles require substantively complex, technical expertise—involving logic, reasoning and numeracy. Examining the role of these specific types of skills contributes to the literature assessing the role of skills type in determining cognitive ageing, and provides a useful complement to the occupation type measures described above.

The measure of STEM-skill usage, henceforth termed *technical skill*, uses variables from the Abilities and Skills from the matched O*NET data (the questions in the Ability and Skill domains were filled out by external job analysts, rather than the job incumbents themselves). This measure was matched to the occupation of first job after leaving school.[2] The Abilities items used are: deductive reasoning; information ordering; mathematical reasoning; number facility. The Skills items are: mathematics; science; technology design; programming. For each item, there is a variable rating the level of each decriptor used in the job, and its importance. The average of the level and importance variables across the Skills and Abilities items identified above are used as a simple continuous measure, where a higher value indicates greater usage of these skills—on average—in an occupation.

Analytical sample

All three subsamples of the US were employed: the General Population Sample, continuing BHPS participants, and the Ethnic Minority Boost sample. Survey weights were employed throughout the analyses, which weights adjust for unit non-response—the combined probabilities of being selected into the BHPS, GPS and EMB and continuing to Wave 3 of the survey—and the complex survey design.

Since the location of school was not available among the cohorts considered, it was not

[2] At the time of writing, the current occupation variable in the US was coded as SOC2000. Only when respondents switched to a new job, the new job was coded as SOC2010. Therefore it was not possible to match the O*NET measure to complete data on current occupation.

possible to ascertain whether respondents completed their schooling in England or Wales, rather than Scotland or Northern Ireland. In order to restrict the sample to individuals who completed their schooling in areas exposed to the reform as best as possible, those who were born in England or Wales were selected for analyses. Observations which had missing or unavailable information on place of birth were assumed not to be born in England or Wales.

The preliminary descriptive statistics use a sample of respondents born with a window of 5 years either side of date determining exposure to the school reform (01, September, 1957). The sample for the OLS and RD estimation uses a smaller window, determined by an optimal bandwidth-selection procedure.

1.4 Empirical approach

1.4.1 Regression Discontinuity Design

A Fuzzy Regression Discontinuity (FRD) design (Imbens & Lemieux, 2008) was employed to exploit variation in schooling induced by the reform. RD is predicated on treatment, for individuals $i = 1, .., N$, being assigned by the value of a continuous covariate, R_i, falling on either side of a fixed cutoff c. In this application c is the pivotal birth cohort of 01, September, 1957. This cutoff induces a discontinuity in the conditional probability of receiving the treatment given R_i—but not necessarily a jump from 0 to 1. Let Z_i denote exposure to the reform, $(Z_i = 1[R_i \geq c])$, where $1(.)$ is the indicator function, and the treatment is an additional year of schooling. Let $Y_i(1)$ and $Y_i(0)$ denote the potential cognitive outcomes experienced in the presence and absence of the treatment respectively. The FRD estimand of interest is the following:

$$\tau_{FRD} = \frac{E\left[\,Y_i(1)\,|\,R_i = c\,\right] - E\left[\,Y_i(0)\,|\,R_i = c\,\right]}{E\left[\,D_i(1)\,|\,R_i = c\,\right] - E\left[\,D_i(0)\,|\,R_i = c\,\right]} \tag{1.1}$$

Assuming the reform only changes behaviour in one direction (monotonicity), then the FRD estimator yields an average treatment effect at the cut-off among the sub-population of compliers (i.e., those who were causally induced to take an extra year of schooling, which otherwise they would not have taken). In this case, the FRD treatment effect is the ratio of the sharp RD effect and the average effect of the reform on treatment, as in Equation (1.2) (Hahn *et al.* , 2001).

$$\tau_{FRD} = \frac{\lim_{r\downarrow c} E\left[\, Y_i \mid R_i = c \,\right] - \lim_{r\uparrow c} E\left[\, Y_i \mid R_i = c \,\right]}{\lim_{r\downarrow c} E\left[\, D_i \mid R_i = c \,\right] - \lim_{r\uparrow c} E\left[\, D_i \mid R_i = c \,\right]} \tag{1.2}$$

This framework also extends to the case of a multi-valued measure of years of schooling, in which the compliers are defined as those induced to take at least d years of schooling when otherwise they would have taken fewer than d (Angrist & Imbens, 1995)). In that case, the effects computed are an average of the per-year treatment effects associated with each additional year of schooling, weighted by the proportions of compliers at each level of schooling. However, in this application there is little weight placed on the schooling levels above 16—the treatment effect is mainly informed by differences in outcomes on the 15 to 16 years margin. The main analyses presented in this paper are for a binary treatment, equal to zero if pupils left school at age 15 years, and one if they left at 16 years of age, and missing for those who stayed on to 17 years or older. This is also to facilitate the computation and interpretation of the analyses of mechanisms in the subsequent section, which is more straightforward with a binary treatment. Results with a continuous treatment are presented in Appendix 2.E.

Using a uniform kernel, and the same bandwidth h for the outcome and treatment equations, leads to a numerical equivalence between the FRD estimator and the Two-Stage Least Squares (2SLS) estimator (Hahn *et al.* , 2001). Taking this approach, the parameters of the equations described in Equations (1.3) and (1.4) were estimated using 2SLS. Let Y_i denote the cognitive outcome observed for individual i. $f(R_i - c)$ comprises the centred running variable, interacted with the reform dummy. The vector X_i includes pre-treatment covariates, and u_i and v_i are idiosyncratic errors.

The estimating equations are described as follows:

$$Y_i = a_o + a_1 \hat{D}_i + f(R_i - c) + X_i' a + u_i \tag{1.3}$$

$$D_i = \gamma_o + \gamma_1 Z_i + g(R_i - c) + X_i'\gamma + v_i \tag{1.4}$$

To select a data-driven optimal bandwidth, the implementation of the MSE-optimal bandwidth developed in Calonico *et al.* (2016) was employed, which accounts for the fuzzy design and clustering of the data due to the discrete running variable (Bartalotti & Brummet, 2016). This was combined with a sensitivity analyses through a consideration of a range of bandwidths. The running variable is month-year of birth. This allows a comparison of units very close to the treatment cut-off. However this measurement is still a discretisation compared, as a more granular measurement of assignment (e.g., day-of-birth) was not available in the data. The standard errors were clustered by month-year of birth to account for group-level variation induced by the discrete nature of the running variable. This relies on a model in which the fitted function, through the discrete running variable, approximates the true continuous function, and the consequent specification errors are random and identical (Lee & Card, 2008; Cattaneo *et al.* , 2017a).[3]

In very small sample windows around the reform date it is more credible that the reform, as a local instrumental variable, is plausibly exogenous without conditioning on further covariates. Other discontinuities exactly coincident with the RoSLA are unlikely. In this case, the purpose of including the covariates is to increase the precision of the estimates by reducing residual variation. In larger sample windows, the concern remains that there may be confounding of education level and cognitive function based on unobservable functions of birth cohort. For example later cohorts, exposed to the reform, may have experienced more favourable conditions in early childhood, and have more educated parents, than the earlier pre-reform cohorts. Although these trends are unlikely to be discontinuous at the treatment cut-off, they may still be picked up in larger sample windows. Therefore, this concern motivates adjustment for potential confounding variables which may capture any such effects. The covariates included in the RD models presented in Table 1.3 are: a quadratic term in age, dummy variables indicating gender, interview month, interview

[3] I considered examining a wider sample window (5 years either side of the cut-off), in conjunction with a quadratic trend in the running variable. However, higher order polynomials are less reliable for the estimation of RD treatment effects—due to over-fitting or biases at boundary points - therefore I restricted attention to the local linear and quadratic cases (Gelman & Zelizer, 2015; Cattaneo *et al.* , 2017b,a).

year (because the US survey waves span approximately two years).

In further sensitivity analyses (Appendix 2.E), month-of-birth dummies were added as covariates. The rationale for considering month-of-birth is to capture any seasonality effects: systematic variation in month-of-birth by family background which could also be related to later education, health and cognitive outcomes (Buckles & Hungerman, 2013; Crawford *et al.* , 2011). Additionally, since the implementation of the 1972 reform coincides with the start of the school term, exposure to the reform will coincide with any age-in-grade effects, and therefore be correlated with schooling and, potentially, cognitive outcomes.

1.5 Unpacking the black box using mediation analysis

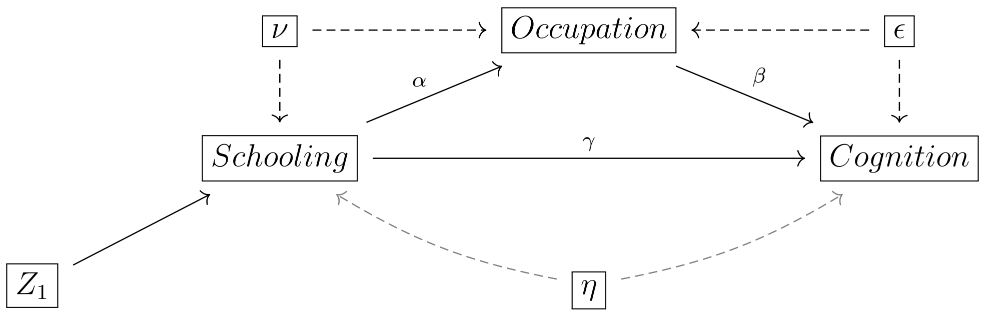
Mediation analysis is one way to quantitatively assess the pathways through which a cause affects an outcome. Mediation analyses offers a principled approach to test economic (and other social science) theory without employing a full structural model. A key motivation is to understanding *how* a policy intervention, or RCT, works (e.g., see Heckman & Pinto (2015) for an application of mediation to an RT in economics). Although randomised studies provide an internally valid estimate of the treatment effect, their underlying workings often remain a black box. Analysis of mechanisms can refine the workings of an intervention. Can we better target the primary mechanism, or eliminate costly and ineffective components of an intervention? We might also want to understanding why an intervention failed.

1.5.1 "Traditional" regression-based approach

Mediation analyses has its origins in the early structural economics literature—Haavelmo (1943) and Goldberger (1972)—however until recently, most studies in the social sciences followed the approach outlined in a seminal article by Baron & Kenny (1986), who defined mediation analysis as *"generative mechanisms through which the focal independent variable is able to influence the dependent variable of interest"*.

The Baron & Kenny (1986) approach can be summarised as follows. Let D_i denote the observed binary treatment status of individual i, M_i the observed mediator value, Y_i the observed outcome. X_i represents one or more covariates which may be thought to influence the other variables. D_i occurs first, then M_i and finally Y_i. For now, assume both the mediator and outcome are continuous random variables, and that X_i are observed pre-treatment. We are interested in the contribution of mediator M_i to the total effect of D_i on Y_i—the 'indirect effect'—versus the contribution of all other unmodeled mechanisms (the 'direct effect'). Figure 1.1 depicts a stylised form of the pathways of interest. The dashed lines represent potential unobserved confounding.

Figure 1.1: Directed Acyclic Graph of education, cognitive outcomes and mediator



Notes: This figure shows a stylised depiction of the association between schooling, occupation and cognition. The dashed arrows indicate the potential confounding effects of unobserved variables ν, η, ϵ. α, β and γ denote estimates of the partial effect of schooling on occupation, occupation on cognitive outcomes, and the direct effect of schooling on cognition, net of occupation, respectively. Z_i denotes an instrumental variable which affects schooling, with no direct effect on occupation or the cognitive outcomes.

Given certain (stringent) assumptions, a linear-regression based approach can be used to retrieve the direct and indirect effects (Huber *et al.* , 2016). An estimate of the indirect (mediated) effect can be obtained using the so-called *product of coefficients* method, multiplying together the partial of effect of the treatment on the mediator and the partial effect of the treatment on the outcome: $\hat{\alpha} \times \hat{\beta}$ in Equations 1.5 and 1.6.

$$E[M_i|X_i, D_i] = \mu_1 + \alpha D_i + X_i'\phi_1 \tag{1.5}$$

$$E[Y_i|X_i, D_i, M_i] = \mu_2 + \gamma D_i + \beta M_i + X_i'\phi_2 \tag{1.6}$$

Alternatively, the *"difference in coefficients"* method first computes the total effect from a the regression of Y_i on D_i (Equations 1.7). Second, it adds M_i as a covariate (Equation 1.6).

$$E[Y_i|X_i, D_i] = \mu_o + \tau D_i + X_i'\phi_o \tag{1.7}$$

The reduction in the effect of the treatment after controlling for a mediator is often interpreted as the proportion of the treatment explained by the mediator. To see this, note that the total effect, τ_i, comprises the sum of the direct and indirect effect, $(\alpha \times \beta) + \gamma$.

Comparing the coefficient on the treatment dummy in Equation 1.7, with the coefficient on the treatment variable after controlling for the mediator, Equations 1.6, is equivalent to comparing the difference between the estimate γ and τ. Since $\tau = (\alpha \times \beta) + \gamma)$, the difference between the γ and τ is $\alpha \times \beta$: the so-called "explained" or mediated effect.

This framework motivates the commonly used approach of adding post-treatment variables to a model as "controls", and observing the extent to which the treatment effect coefficient is reduced. An example is the gender wage gap, where a common approach is to add post-treatment variables, such as occupation, to a regression of wages on gender—to assess to what extent they "explain" the wage gap. Huber *et al.* (2016) shows that this type of decomposition of gender or ethnic wage gaps, into explained and unexplained components, is generally invalid—it is equivalent to assuming a overly simple mediation model, and lacks causal interpretability.

Indeed, the methods described above require strong assumptions to be valid: (i) the errors terms should be uncorrelated across equations, implying no unmeasured confounding of the treatment, mediator and outcome; (ii) the conditional expectations of the mediator and outcomes should be linear and additive; (iii) the effects of the mediator cannot interact with treatment status.

1.5.2 Occupation as a mechanism

The FRD analyses yields a total treatment effect (among compliers). Since any effects of schooling on cognitive function could operate either directly or indirectly, though subsequent opportunities and choices that schooling affords, this section tests a number of hypothesised underlying mechanisms.

As described in the previous section, one approach to answering this question is to employ a so-called mediation analysis. Mediation analysis aims to empirically disentangle the causal mechanisms through which a treatment exerts its effect on the outcome. A mediator is a variable which lies on the causal pathway between treatment and outcome. In this framework, a total effect may be decomposed into two components: an indirect effect, operating through the mediator of interest, and a direct effect operating through all other unmodeled intermediate variables. Recent literature has clarified the assumptions under which these mechanisms are identified (see, for instance, Imai *et al.* (2011); Huber *et al.* (2016); Keele *et al.* (2015) for recent reviews and applications).

In contrast to the older literature, which employed more restrictive parametric methods (Baron & Kenny, 1986), the recent focus has been on non-parametric identification, leading to the development of a range of flexible estimation approaches.

This paper employs the results developed in Yamamoto (2014) and Park & Kürüm (2018), who shows that—-given certain assumptions—the LATE can be decomposed into components explained, and unexplained, by a mediator of interest. To do so, Yamamoto (2014) and Park & Kürüm (2018) develop a proof of non-parametric identification of average causal mediation effects among compliers, and propose a flexible estimation approach to implement the method. This section draws on that paper to briefly outline the framework.

Let $D_i \in \{0, 1\}$ denote the level of treatment selected by individual i, where $i = 1, ..., N$. $D_i = 1$ if a pupil leaves school at 16, and $D_i = 0$ if a pupil leaves at 15 years. Z_i denotes the observed value of the instrument. $D_i(Z_i)$ indicates the potential treatment state for individual i; the treatment they would select depending on the value of the instrument. The observed schooling level can then written as $D_i = D_i(Z_i) = Z_i D_i(1) + (1 - Z_i) D_i(0)$. Let M_i and Y_i denote observed mediator and observed cognitive outcomes, respectively, and X_i a vector of observed pre-treatment covariates. $M_i(d)$ indicates the potential mediator state, depending on the level of schooling chosen. $M_i(1)$ is the value of the mediator chosen under treatment level 1, and $M_i(0)$ is the value the mediator would take under treatment level 0. The potential cognitive outcomes are denoted $Y_i(d, m)$, depending on treatment and mediator.

The local average treatment effect, incorporating the choice of mediating variable, can be written as follows:

$$\tau = E\,[\,Y_i(\,1, M_i(1)\,) - Y_i(\,0, M_i(0)\,)\,|\,D_i(1) = 1,\ D_i(0) = 0\,] \tag{1.8}$$

Varying the treatment exogenously, but fixing the mediator at the its potential value for

$d \in \{0, 1\}$ yields the average direct effect among compliers: the Local Average Natural Direct Effect (*LANDE*).

$$\zeta(d) = E\,[\,Y_i(\,1, M_i(d)\,) - Y(\,0, M_i(d)\,)\,|\,D_i(1) = 1,\; D_i(0) = 0\,] \tag{1.9}$$

Fixing the treatment at $d \in \{0, 1\}$, but varying the mediator to its potential values under treatment and non-treatment yields the average indirect effect among compliers: Local Average Complier Mediated Effect (*LACME*).

$$\delta(d) = E\,[\,Y_i(\,d, M_i(1)\,) - Y_i(\,d, M_i(0)\,)\,|\,D_i(1) = 1,\; D_i(0) = 0\,] \tag{1.10}$$

The LACME and LANDE, defined on opposing treatment states, sum to the LATE.[4] $\delta(1)$, for example, represents the difference in two potential outcomes: $Y_i(1, M_i(1))$ represents the observed cognitive outcome if pupil i stays at school until 16 years; $Y_i(1, M_i(0))$ represents the cognitive outcome under a counterfactual scenario, in which the pupil again stays at school until 16 years, but then selects the occupation they would have chosen if they had left at 15 years. The difference between these two potential outcomes is the effect of the change in the mediator induced by the treatment, fixing the direct effect of the treatment.

Conversely, $\zeta(1)$ represents the difference in potential cognitive outcomes between leaving school at 15 years and 16 years, holding occupation constant at the level which would be chosen after leaving at 16 years. Therefore, this is the portion of the treatment effect not transmitted through the mediator. The notation $\zeta(d)$ and $\delta(d)$ suggest the possibility of interaction between mediator and treatment.[5]

[4] The following should be conditioned on being a complier; notation omitted for brevity.

$$\begin{aligned} LATE &= E\,[\,Y_i(\,1, M_i(1)\,) - Y_i(\,0, M_i(0)\,)] \\ &= E\,[\,Y_i(\,1, M_i(0)\,) - Y(\,0, M_i(0)\,)] + E\,[\,Y_i(\,1, M_i(1)\,) - Y(\,1, M_i(0)\,)] = \zeta(0) + \delta(1) \\ &= E\,[\,Y_i(\,1, M_i(1)\,) - Y(\,0, M_i(1)\,)] + E\,[\,Y_i(\,0, M_i(1)\,) - Y(\,0, M_i(1)\,)] = \zeta(1) + \delta(0) \end{aligned}$$

[5] For example, this would mean that the indirect effect—the effect of schooling-induced occupation choice on average cognitive outcomes—could differ between those who leave at 15 and those leave at 16 years.

$\zeta(d)$ and $\delta(d)$ are counterfactual quantities, since we do not observe any individual with, for instance, the value of the mediator they would have selected under the counterfactual treatment. Given the counterfactual nature of the effects of interest, identification of causal mechanisms requires accounting for potential endogeneity of not only the treatment, but also the mediating variable.

The assumptions required to identify the LANDE and LACME include, first, an exclusion restriction for the instrument, and second, monotonicity—as required in a standard instrumental variable analysis. The third assumption is the so-called *local sequential ignorability* assumption (described in Equations 1.11 and 1.12). The first component of this local sequential ignorability assumption, described Equation 1.12, requires conditional independence of the instrument with respect to both the potential cognitive outcomes and potential mediator states. This is satisfied by a valid instrument for schooling. The second component, Equations 1.11, requires the *mediator* to be conditionally independent of the potential cognitive outcomes, given the treatment (years of schooling) and pre-treatment covariates, among the compliant subpopulation (clearly this is less restrictive than requiring this independence across all compliance types).

$$Y_i(d,m),\ M_i(d'),\ D_i(z)\ \perp Z_i \mid X_i \tag{1.11}$$

$$Y_i(d',m)\ \perp M_i(d) \mid D_i = d, X_i,\ \textit{type} = \textit{complier} \tag{1.12}$$

In this application, this final condition will be satisfied if we believe that there is no unobserved variable systematically influencing both occupation choice and cognitive function, within cells defined by both education level and the pre-treatment covariates, among the compliers to the reform. In contrast to the choice of education level, it is more difficult to think of specific confounding variables which would play this role. One candidate is parental occupation, and sensitivity analyses are conducted to the addition of this potential confounder[6]. Additional maintained assumptions include the absence of any confounders

[6] In Appendix 1.G, I conduct a sensitivity analysis which assesses the impact of deviations from conditional independence

of the mediator and outcome which are themselves caused by the treatment (so-called *intermediate confounders*).

An alternative option to handle endogeneity of mediator is to deploy a second instrumental variable. For example, Dippel *et al.* (2017) propose a method which uses one instrument to handle the endogeneity of both the treatment and mediator, to investigate the role of labour market impacts in explaining trade liberalisation's effects on political polarisation [7]. Tchetgen *et al.* (2012) and Huber *et al.* (2017a) have considered non-parametric identification using two instruments, which also requires an instrument for the mediator as well the treatment. In the current application, I have explored a number of potential instruments (local labour market conditions, parental occupation), but failed to identify a credible instrumental variable for occupation choice.

The estimator employed for the LANDE and LACME is described here in Appendix 1.F. It is expressed as functions of conditional expectations (and densities) of the outcome, treatment and mediator. The estimation procedure involves fitting flexible predictive models to the treatment (years of schooling), mediator and outcome. The relevant conditional expectations, or conditional densities, are then computed and plugged into the respective formulas (in Appendix 1.E). Standard errors were obtained via bootstrapping, with 1000 replications.

One caveat of the analyses presented here is that each of the intermediate variables are treated as independent from each other: the LANDE and LACME are computed separately for each mediator. This does not allow for causal dependence between the mediators. Conceptually, the purpose is to examine these variables as alternative measures of the same potential channel, rather than the intention being to model them jointly as a system and disentangle the distinct role of each mediator. Extending the framework to multiple dependent mediators would require more complex sets of potential outcomes: defining the potential mediator values in response to the choice of the other mediators, and the imposition of further assumptions about confounding between mediators themselves (e.g., see Park & Kürüm (2018)). This type of extension is left for further work.

[7] I have endeavoured to apply this approach to the current paper, however the assumptions required failed to hold, leading to a weak instrument problem, and I concluded it is not a credible approach for this setting.

1.6 Findings

This section considers the findings from the OLS and RD specifications, sensitivity analyses, and an examination of the channels through which these effects operate.

1.6.1 OLS and RD results

Table 1.3 reports OLS and RD results of the effect of schooling on Word Recall, Numeric Ability, Verbal Fluency and Subtraction, respectively. The sample used in this table has two main restrictions: 1) respondents who left school before the age of 17 years (this group represents the bulk of the compliers), 2) respondents who have complete data for each of the three occupation type measures examined in the analyses of mechanisms (for comparability). Estimates of the effects of schooling on cognitive outcomes for the unrestricted sample are presented in Appendix Table 1.6.

The findings show a positive association between years of schooling and each cognitive outcome. An additional year of schooling is associated with a 0.15 standard deviation increase in Word Recall, 0.18 standard deviation increase in Verbal Fluency, 17 percentage point increase in Numeric Ability and a 10 percentage point increase in Subtraction. These results corroborate those found in other literature: the positive correlation between education and cognitive outcomes persists into later life. However, these findings may not necessarily represent causal effects, due to omitted variable bias. The RD specification addresses this issue by exploiting plausibly exogenous variation in schooling.

Table 1.3 shows the RD estimates using a linear specification in the running variable (month-year of birth cohort). A discontinuity in average years of schooling is present at the reform cut-off. This is reflected in the first-stage results: the average difference in years of schooling between those exposed and unexposed to the reform is between 0.33 and 0.44, depending on the sample window used. The RD estimates exploit this jump in years of schooling. Computed at the optimally-selected bandwidth, 30 month-year of birth cohorts, the RD estimates show that an extra year of schooling is associated with a statistically significant 0.53 standard deviation increase in Word Recall. Considering the two components of the word recall measure separately, i.e., delayed and immediate recall, the effect size is larger for delayed recall (data not shown). For Verbal Fluency, the effect size is 0.15 standard deviations, but not statistically significant at conventional levels. Again these estimates are derived from a small sample window, comprising those born 30 months either side of the threshold.

Although there is a positive association between schooling and the two measures relating to numeracy—Serial 7 Subtraction and Numeric Ability—shown in the OLS estimation results, this is not reflected in the RD specifications. The effect size is similar for Numeric Ability, an additional year of schooling is associated with a 17 percentage point increase the probability of successfully answering 4 or 5 questions out of the 5 numeric ability questions asked. For Serial Subtraction, the sign of the effect switches, with an effect size of minus 14 percentage points. These effects are not statistically significant at conventional levels. This may be due to ceiling effects: clustering toward the top of the distribution of performance on these tests, such that the measure may have little ability to distinguish between moderate and high functioning. Treating these variables as continuous variables, rather than dichotomising them, produces similar results.

Education may affect men and women differently; men and women have both different labour market trajectories and different levels of cognitive function in older ages. Table 1.6 in Appendix 2.E reports results from the same sample and specification used in Table 1.3, now separately by gender. At the optimal bandwidth, no statistical evidence is detected for a differential association between schooling and the cognitive outcomes by gender. The size of the treatment effect by gender does vary by bandwidth and the sample definition, however the difference in estimates between men and women in other sample windows is also not statistically significant. Given the small sample sizes in the gender subgroups, a more complete examination of gender differences is not pursued here.

RDD sensitivity analyses

An important choice is the choice of sample window (the bandwidth). To assess the sensitivity of the findings to a range of bandwidths, the point estimates and confidence intervals by bandwidth were examined. For instance, Figure 1.4 plots the treatment effect on each cognitive outcome, and 99%, 95% and 90% confidence intervals, for a range of bandwidths. The optimally-selected bandwidths are indicated by vertical lines. There is variation in the effect size by bandwidth: for Word Recall, the treatment effect ranges from a minimum of 0.24 to a maximum of 0.74 across bandwidths from two to five years. This shows that the estimates do vary by choice of sample, however they remain positive and significant at least the 10% level across a reasonable range of bandwidths. This is in contrast with the other outcomes, where the effects size oscillates around zero and no statistically significant effects are detected across all values of bandwidth choice.

These plots are presented again in Appendix 2.E, Figure 1.7, using a local quadratic, rather than local linear, RD specification: the local quadratic specification has the running variable included as a quadratic, rather than linear, term. These results corroborate those from the simple linear case, and show the results are robust to a more flexible specification of the birth cohort trends. Appendix Table 1.6 and Appendix Figure 1.8 reports results—the causal effect of years of schooling on each cognitive outcome—for the full sample with no restrictions (i.e., including those who left school at ages 17, 18, or 19 years, and including those who did not report data for the occupation variables). The main findings are qualitatively similar to those in the restricted sample. Figure 1.6 also reports on a specification with month-of-birth dummies as covariates, demonstrating that the results are similar with this addition. Overall, the size of the estimates of the effects of schooling on cognitive outcomes do vary with the sample definition, covariate choice and bandwidth, and are not statistically significant in all specifications examined. However, taken as a whole, the positive, statistically significant, effects across a range of reasonable bandwidths and sample definitions provides a strong case for the veracity of the main results.

The corresponding findings in the most similar paper—Banks & Mazzonna (2012), who exploited the 1947 raising of the school leaving age from 14 to 15 year—include a causal effect of an additional year of schooling on Word Recall, which ranges between about one-fifth of a standard deviation to two-thirds of a standard deviation among men, and between one-fifth of a standard deviation to two-thirds of a standard deviation among women. Their preferred estimates are an effect of half a standard deviation among men, and 0.4 among women. For Verbal Fluency, the effect sizes are similar to that for Word Recall among men, and statistically significant at the 10% level. Among women, no evidence was detected for any effects on Verbal Fluency—the effects oscillated around zero and were not statistically significant.

1.6.2 Channels

The RD results revealed a positive causal effect of schooling on Word Recall. This section reports on the analyses of potential underlying mechanisms.[8]

Table 1.4 presents the results of the analysis of causal mechanisms. The table presents the LATE, the LACME (portion of the LATE explained by the mediator) and LANDE (remaining portion of the LATE, explained by all other unmodeled intermediate variables). These quantities are calculated separately to test the role of three alternative mediator variables—first occupation type, current occupation type, technical skills—presented across the three columns.

The indirect effects among compliers are presented separately depending on whether the treatment status is fixed at the control or treatment status (leaving school at 15 years or 16 years). This allows for an interaction between treatment and mediator; the size of the indirect effect can differ by treatment level. The table shows that occupation type of first job explains 0.0% of the LATE for the controls, and 0.02 (3.8%) for treatment, although these are estimates are very imprecise and not significant at conventional levels. For the occupation type of current job, these figures are both 0.06 (11.3%). Considering the level of technical skills used in the first job, the size of the indirect effect is negligible at 0.00 for the controls, and 0.10 (18% of the total LATE) for treatment. These models were fit to the sample window defined by the optimally-selected bandwidth of 30 month-year of birth cohorts. The size of the estimates vary as the sample window is modified, however the qualitative conclusions remain the same. The findings are also robust to the addition of parental occupation as an additional covariates (see Appendix 1.F, Table 1.9).

[8] Further analyses in Appendix 2.E.1 presents simple estimates of the effects of schooling on occupation.

In Appendix 1.G, further sensitivity analyses are presented to assess the impact of deviations from conditional independence. This analysis shows that the findings are robust to moderate deviations from the conditional independence assumption.

The small, positive, indirect effects suggest occupation choice may play a role, consistent with theory—however the estimates of the indirect effect are very imprecisely estimated. This is to some extent due to the nature of the estimation method, which is less efficient than a parametric approach, as can be seen in comparing the precision of the LATE estimated by 2SLS with the nonparametric method. In terms of effect size, the magnitude of the portion explained by occupation is in line with that in other studies. Recent work from the United States found that occupational complexity of the longest held job explained between 11% and 22% of education's association with cognitive function, in a sample of adults aged over 44 years (Fujishiro *et al.* , 2017). This study was wholly based on a selection-on-observables identification strategy; the findings are less plausible as causal estimates.

Fujishiro *et al.* (2017) examined mediation effects among gender and race subgroups, as well across different margins of education. Occupational complexity explained differing amounts of the effect of education on cognition across subgroups—especially at the highest levels of education. A potential explanation suggested is differential skill mismatch in the labour market: if workers do not find roles which match their education level, they may not obtain the full cognitive returns to their education. Altenatively, for the same occupation, there may be heterogeneous cognitive returns across groups. Although the sample size in the present paper is too small to investigate more granular subgroups, exploring differences in the role of occupation by subgroup—gender, age, ethnicity and region—may represent a useful avenue for future work.

If occupation does represent a causal pathway between education and cognitive outcomes, this suggests a second policy lever after full-time education has been completed—supporting young people into employment, providing training to assist with continued work and engagement in later life, and matching employees with work commensurate with their skill set. Although the effects of occupation may themselves operate through subsequent mediating variables that may come with higher status occupations—autonomy and control over job tasks, positive peer group effects, a healthy and safe work environment—occupational mismatch and employment conditions may still represent a “catch-all" indicator for policy intervention. The positive contribution of technical skill usage provides support that effect is not entirely due occupational social status.

The findings also highlight the role of the type of work in sustaining mental performance. Continuing to work into later life in a role which is not engaging, or is physically hazardous, would not be expected to provide continuing benefits. In this case, work-like activities which provide mental engagement, such as volunteering or other non-market activity, may be effective in maintaining long-term cognitive health. Empirical evidence for this hypothesis is provided in Andel *et al.* (2015), who assess the role of mid-life occupational complexity and leisure activity on late-life cognition, and suggest that such non-market activity can compensate for a lack challenging paid work, and vice versa.

1.7 Concluding Remarks

Continued increases of the minimum school leaving age aim to improve the educational, economic and social prospects of those individuals who would otherwise choose to drop out early. Successive changes of this kind have increased average years of education over the the last century. Given these policies are not without cost, the size of social and private returns remains an important question. This paper used a change in compulsory schooling laws, enacted in 1972 in England and Wales, to study the effects of schooling on later-life cognitive performance. The findings show that an additional year of high school confers a protective effect on memory, ranging from one- to two-thirds of a standard deviation, depending on the sample and model specification employed. Little evidence was detected for effects on numeric ability or verbal fluency; the effect sizes were generally close to zero, and statistically insignificant.

These results are in line with previous studies which have generally found a large impact of schooling on working memory, across a range of time periods, countries and estimation strategies, among those at the lower end of the schooling distribution (Glymour *et al.* , 2008; Banks & Mazzonna, 2012; Schneeweis *et al.* , 2014; Mazzonna, 2014). The most similar study Banks & Mazzonna (2012), exploited the 1947 raising of the school leaving age in England and Wales. The 1947 reform had large effects on staying-on rates—at a lower margin of schooling than the 1972 reform—inducing about 50% of the affected cohort to remain in school to 15 years, rather than 14 years.

The effects of schooling on memory are within a similar range of magnitudes as in Banks & Mazzonna (2012) (e.g., effects of 0.5 standard deviations among men, and 0.4 standard deviation among women): evidently, the cognitive returns to basic education have not been exhausted, for an outcome which is especially relevant for the onset of cognitive impairment. In contrast to memory, the present study does not detect any causal effect on Verbal Fluency or the numeric ability measures. This finding may be due to diminishing returns to years of education, to lack of sensitivity of the cognitive battery measures employed, or the fact that the sample examined in Banks & Mazzonna (2012) were older, by about a decade—the full effects on cognitive outcomes may not materialise until older ages. The absence of evidence for an effect on simple measures of numeric ability is consistent with other studies using changes in compulsory schooling laws, which have failed to detected effects of education on both numeric ability, measured both by simple cognitive battery tests (Schneeweis *et al.* , 2014) and through measures of financial decision-making quality (Banks *et al.* , 2018).

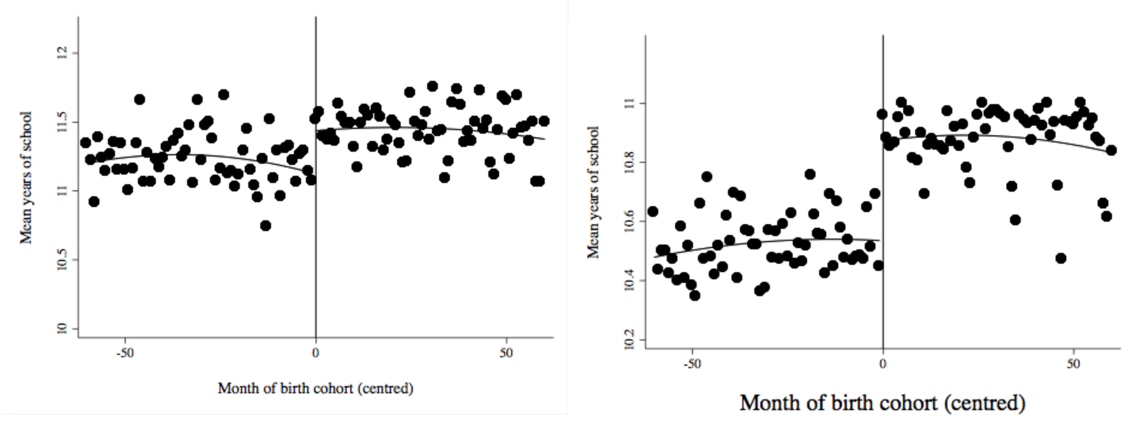
This paper also conducted a formal analyses of the mechanisms shaping the effects of schooling on memory, focussing on the role of occupation type. Occupation type can explain up to about one-fifth of the total effect—in line with other studies (Fujishiro *et al.* , 2017). However, the figures are imprecisely estimated and not statistically significant. Based on the effect sizes, intervening on occupational mismatch and working conditions may exert a small to moderate effect on cognitive outcomes—however given the large standards errors the analyses remain inconclusive.

As global populations age, understanding the drivers of health and functioning of older people is increasingly important. The findings of this paper show that one additional year secondary schooling has a causal effect on memory, an important and policy-relevant component of cognitive functioning. The expansion of public schooling throughout the twentieth century may reduce growth in the burden of adverse cognitive outcomes as the cohorts exposed these reforms age.

Figure 1.2: Years of Schooling by month-year of birth

(a) Full sample

(b) Restricted to those who left school before age 17 years

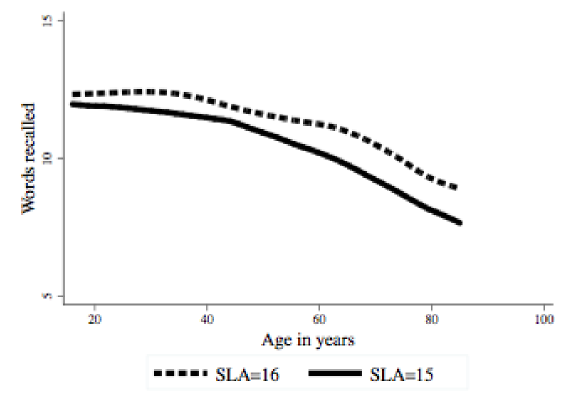


Notes: Sample means of years of school by month-of-birth cohort.

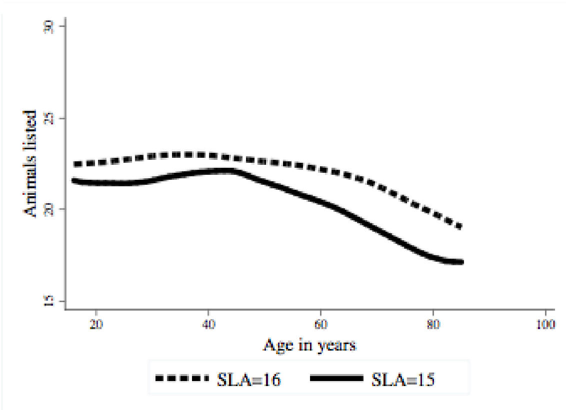
Figure 1.3: The evolution of cognitive outcomes by age

(a) Word Recall

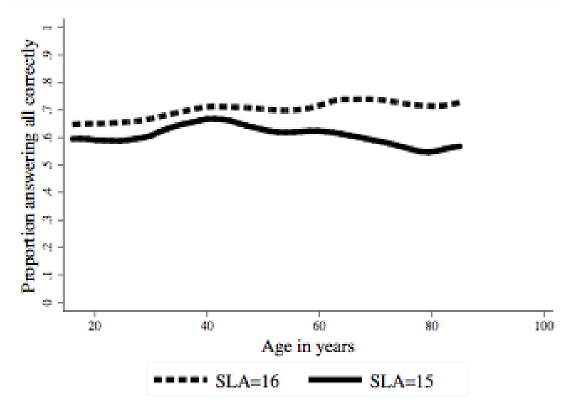
1



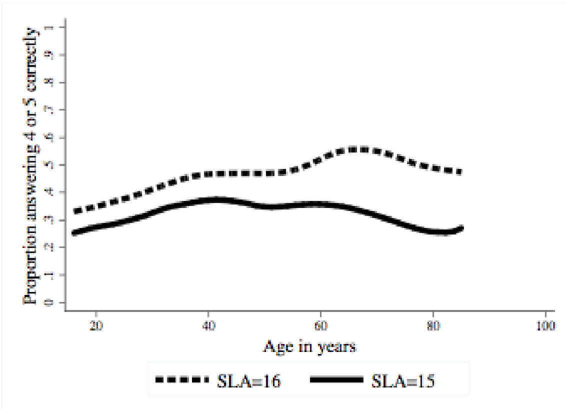
(b) Verbal Fluency



(c) Serial Subtraction



(d) Numeric Ability

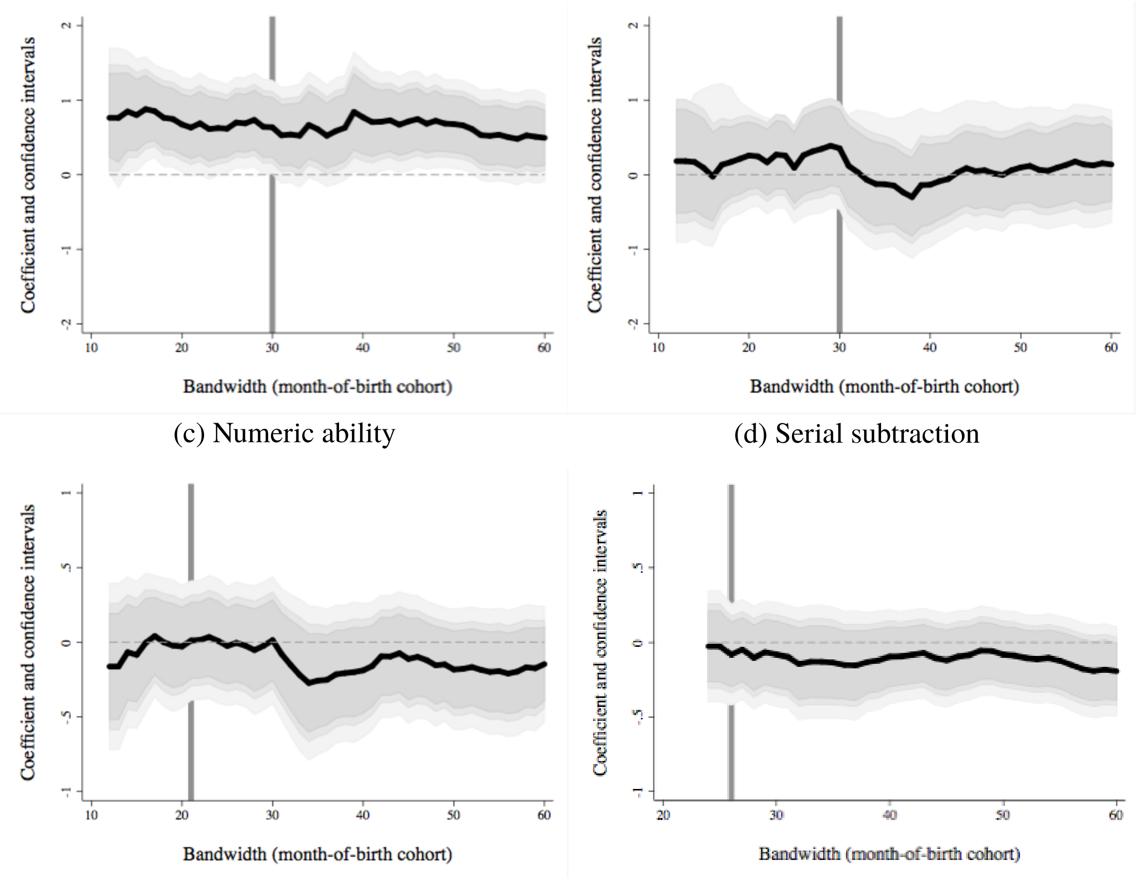


Notes: Mean cognitive outcomes (Word Recall, Verbal Fluency, Serial Subtraction and Numeric Ability) by age-at-interview and school leaving age (SLA), among those who left school at either 15 years or 16 years of age.

Figure 1.4: Sensitivity analyses

(a) Word Recall (b) Verbal fluency

(c) Numeric ability (d) Serial subtraction



Notes: The bold black line plots coefficients from a fuzzy Regression Discontinuity Design assessing the effect of years of school on each cognitive outcome (Word Recall, Verbal Fluency, Numeric Ability, Serial Subtraction) across a range of bandwidth choices (22 month-year of birth cohorts to 60 month-year of birth cohorts). The grey areas depict 90%, 95% and 99% confidence intervals around the treatment effect. The vertical line indicates the optimally chosen bandwidth.

Table 1.1: Age left school (SLA), Before and After Reform

	Men			Women		
	Treatment status			Treatment status		
SLA	Non-treated %	Treated %	Total %	Non-treated	Treated	Total
15	37.8	7.9	21.7	37.7	9.2	22.0
16	30.1	60.3	46.4	33.1	60.0	47.9
17	9.9	10.4	10.2	9.5	9.9	9.7
18	19.9	19.7	19.8	18.9	19.8	19.4
19	2.3	1.6	1.9	0.9	1.1	1.0
Total	100.0	100.0	100.0	100.0	100.0	100.0

Notes: Survey-weighted column percentages summarising age left secondary school (SLA), for men and women. The column Non-treated is restricted to respondents unexposed to the new the school leaving age, born before 01, September, 1957. The column Treated is restricted to respondents exposed to the new the school leaving age, born on or after 01, September, 1957. The sample is restricted to respondents born within five years of the treatment cut-off.

Table 1.2: Summary statistics for demographic and cognitive variables

Variable	Mean	SD	Min	Max	N
			(a) Pooled sample		
Female	0.52	0.50	0.00	1.00	4,915
Age	53.54	2.94	48.00	60.00	4,915
Years of school	11.33	1.04	10.00	14.00	4,833
Word recall	11.60	3.23	0.00	20.00	4,797
Verbal Fluency	22.93	6.80	0.00	71.00	4,877
Numeracy	0.53	0.50	0.00	1.00	4,864
Serial Subtraction	0.71	0.45	0.00	1.00	4,779

Notes: Weighted summary statistics for analysis variables. Restricted to respondents born five years before or after the 01, September, 1957.

Table 1.3: OLS and RD estimates of the effect of schooling on cognitive outcomes

	Outcome							
	Word recall		Verbal Fluency		Numeric Ability		Subtraction	
	OLS	RD	OLS	RD	OLS	RD	OLS	RD
	(a) RD treatment effect							
Years	0.15	0.53	0.18	0.15	0.16	0.17	0.10	-0.14
Std. Err	0.06	0.28	0.06	0.38	0.04	0.12	0.04	0.17
	(b) First-stage statistics							
First-stage $\hat{\beta}$		0.33		0.34		0.44		0.38
F-statistic		35.35		36.08		41.21		35.86
	(c) Sample size							
Bandwidth	30	30	30	30	21	21	26	26
N	939	939	950	950	671	671	810	810

Notes: $Panel(a)$ reports results from OLS and FRD regressions assessing the effect of an additional year of schooling on each cognitive outcome. Each model adjusts for the following covariates: a linear trend in month-year of birth cohort (interacted with the reform dummy), indicators for gender, interview month, interview year and a quadratic term in age. The standard errors are clustered by month-year of birth cohort. Survey weights adjusting for unit non-response and sample design were used in all specifications. $Panel(b)$ presents the first-stage statistics. First-stage $\hat{\beta}$ is the coefficient on the reform dummy in the first stage regression. F-statistic is the robust F-statistic for the first-stage. $Panel(c)$ shows the sample size. $Bandwidth$ refers to the number of month-year of birth cohorts included in the estimation sample on each side of the treatment cut-off, selected using a data-driven procedure. N is the sample size used in each regression.

Table 1.4: Causal mechanisms: LACME and LANDE

	Occupation (first job)		Occupation (current job)		Technical skills	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
LATE	0.52	(0.38)	0.53	(0.39)	0.53	(0.39)
LANDE						
$\tilde{\zeta}(0)$	0.50	(0.53)	0.47	(0.84)	0.53	(0.73)
$\tilde{\zeta}(1)$	0.52	(0.74)	0.47	(0.46)	0.43	(0.38)
LACME						
$\tilde{\delta}(0)$	0.00	(0.86)	0.06	(0.32)	0.00	(0.52)
$\tilde{\delta}(1)$	0.02	(0.62)	0.06	(0.79)	0.10	(0.30)
Bandwidth	30	30	30	30	30	30
N	939	939	939	939	939	939

Notes: This table shows estimates of the local average treatment effect (LATE) of one additional year of schooling on Word Recall, and decomposes this into components explained and unexplained by three alternative intermediate variable (Occupation type of first job, Occupation type of current job, Technical skills in first job). $\tilde{\zeta}(0)$ and $\tilde{\zeta}(1)$ ($\tilde{\delta}(0)$ and $\tilde{\delta}(1)$) denote estimates of the LACME (LANDE) when the treatment is fixed to either $d = 0$ or $d = 1$, respectively. Each model adjusts for the following covariates: linear trend in month-year of birth cohort interacted with the reform dummy, indicators for gender, interview month, interview year, age. $Bandwidth$ refers to the number of month-year of birth cohorts included in the estimation sample on each side of the treatment cut-off.

Appendix 1.A Distributions of outcome variables

Figure 1.5 displays the distribution of raw data for the main cognitive outcome variables considered. Word recall and verbal fluency are continuous measures, and were subsequently standardised to have mean zero and standard deviation of 1. The results are robust to excluding zero values for word recall and verbal fluency.

Figure 1.5: Histograms of variables

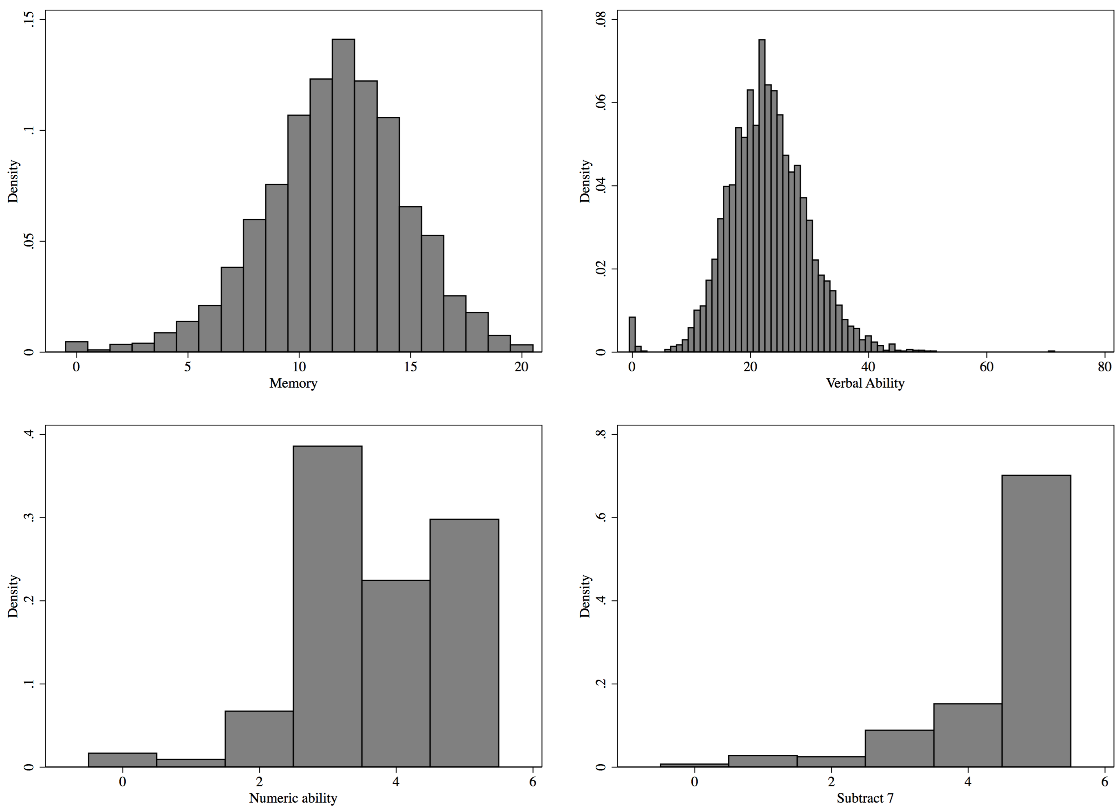


Table 1.5: Summary statistics for occupation variables

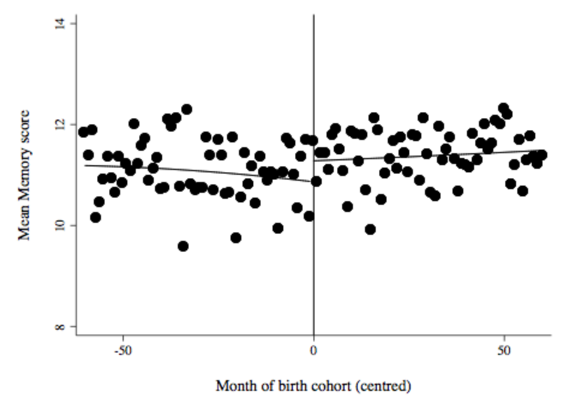
Variable	Column % / Mean	N
NS-SEC (first job)		
Semi-routine; routine; LT unemployed	23.1	777
Lower supervisory and technical	33.7	1,134
Small employers and own account	5.0	168
Intermediate occupations	26.1	876
Lower managerial; Higher managerial and professional	12.0	405
Total	100.0	3,361
NS-SEC (current job)		
Semi-routine; routine; LT unemployed	28.0	1,157
Lower supervisory and technical	7.5	311
Small employers and own account	10.5	437
Intermediate occupations	12.6	520
Lower managerial; Higher managerial and professional	41.4	1,714
Total	100.0	4,139
Skills		
Technical skills (in first job)	16.85	3,389

Notes: Weighted column percentages for occupation variables. Restricted to respondents born five years before or after the 01, September, 1957 who left school before age 17 years.

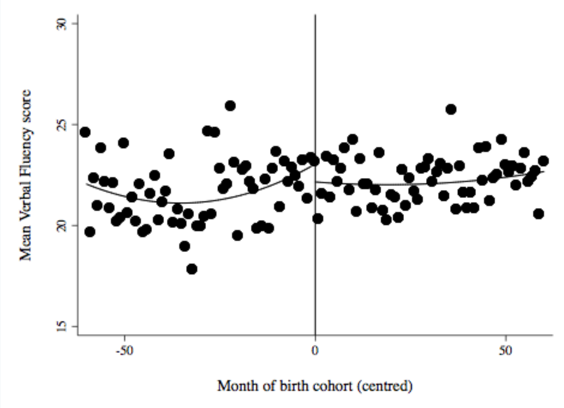
Appendix 1.B Descriptive analyses

Figure 1.6: Cognitive outcomes by birth cohort

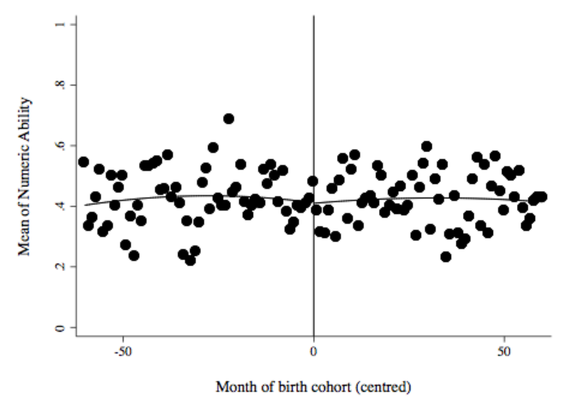
(a) Word Recall



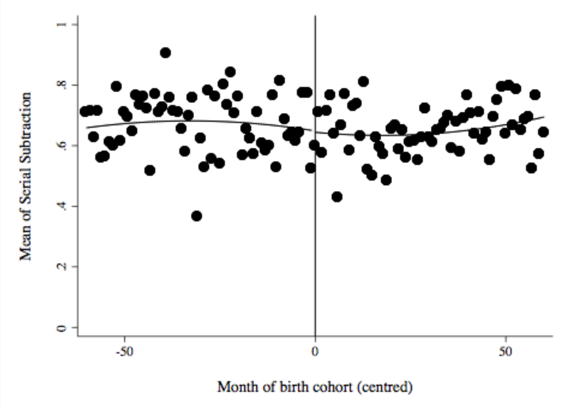
(b) Verbal Fluency



(c) Serial Subtraction



(d) Numeric Ability



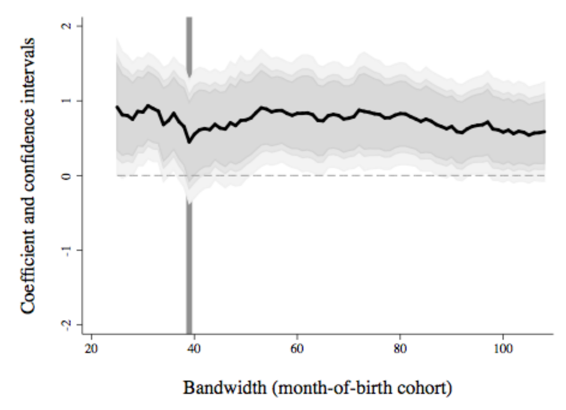
Notes: Sample means of Word Recall, Verbal Fluency, Serial Subtraction and Numeric Ability by month-of-birth cohort, overlaid with a quadratic fit.

Appendix 1.C Sensitivity analyses: Regression Discontinuity Design

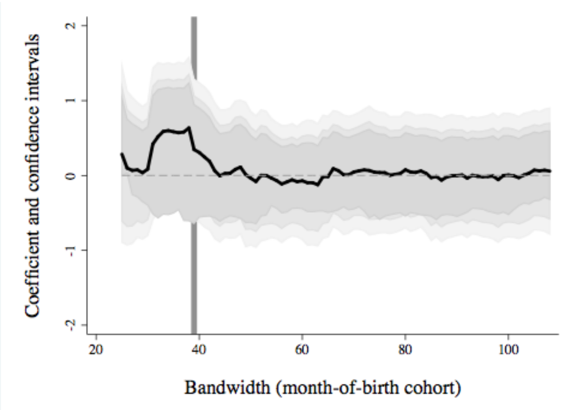
Figure 1.7 assesses the sensitivity of the main results to changes in the bandwidth choice. The treatment effect coefficient and confidence intervals (90%, 95%, 99%) are plotted for models with different bandwidths, from 24 months to 121 months (units born within 2 years either side of the reform up to 10 years either side). As in the main text, this sample is restricted to respondents who left school before age 17 years.

Figure 1.7: Sensitivity analyses

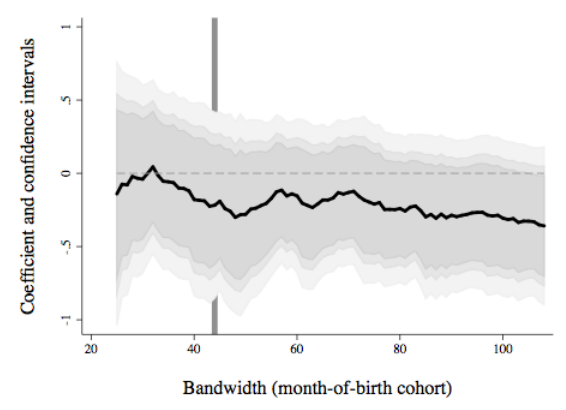
(a) Memory: Quadratic



(b) Verbal: Quadratic



(c) Serial Subtraction: Quadratic



(d) Numeric Ability: Quadratic

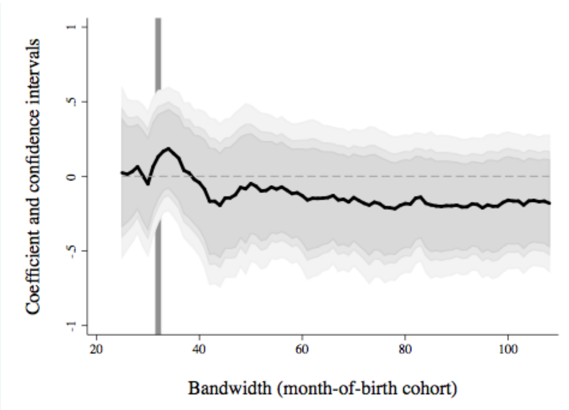


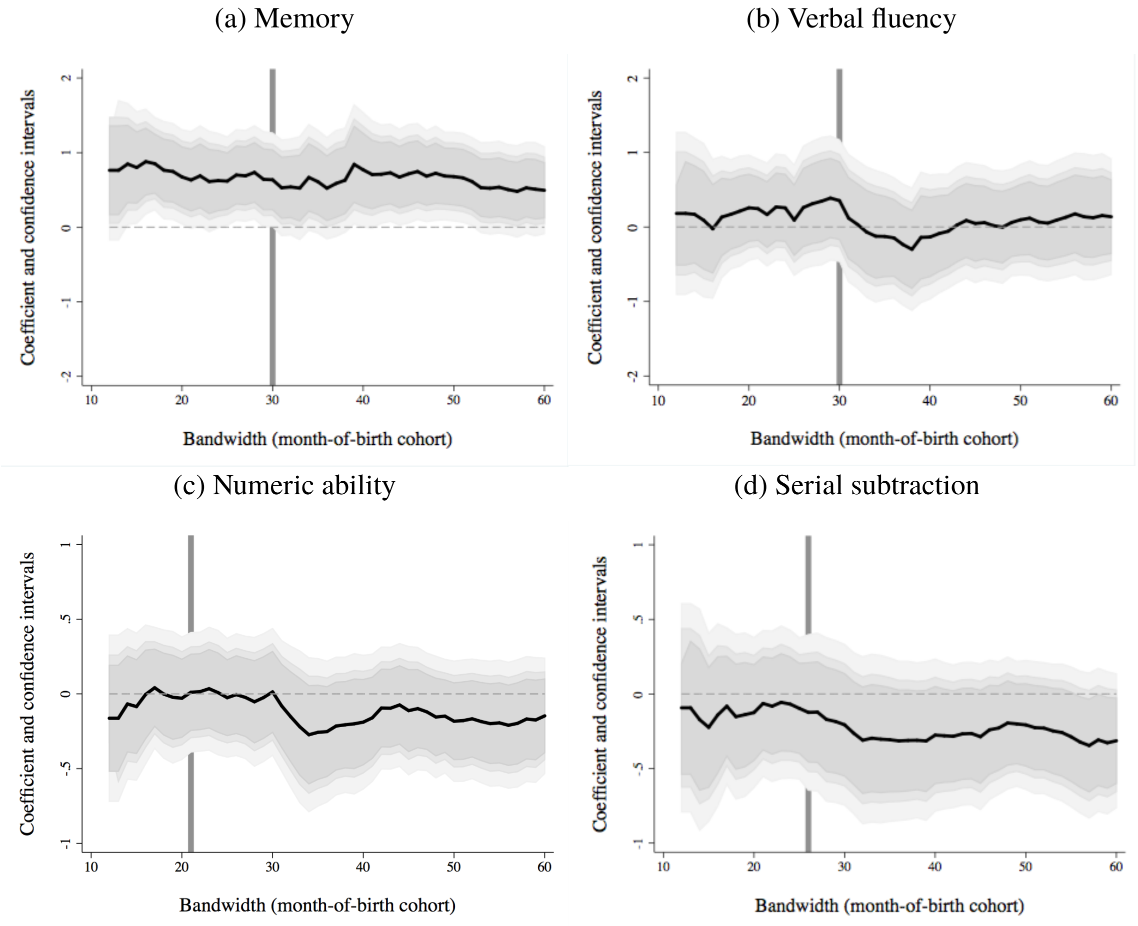
Table 1.6 reports the main results computed for the full sample: i.e., including respondents who left school at age 17, 18 or 19 years, in addition to those who left at 15 or 16 years, and including those who did not report occupation information. Figure 1.8 reports sensitivity analyses to the bandwidth choice.

Table 1.6: OLS and RD estimates of the effect of schooling on cognitive outcomes (full sample, i.e., left school $\leq$ 19 years)

	Outcome							
	Word recall		Verbal Fluency		Numeric Ability		Subtraction	
	RD1	RD2	RD1	RD2	RD1	RD2	RD1	RD2
	(a) RD treatment effect							
Years	0.32	0.41	0.24	0.29	0.01	0.10	0.03	0.02
Std. Err	0.13	0.17	0.19	0.23	0.12	0.16	0.12	0.13
N	2311	2311	2356	2356	1645	1645	2000	2000
	(b) RD treatment effect (Women)							
Years	0.52	0.67	0.53	0.68	0.24	0.59	0.03	-0.07
Std. Err	0.26	0.31	0.34	0.43	0.22	0.71	0.26	0.26
N	1290	1290	1316	1316	916	916	1115	1115
	(c) RD treatment effect (Men)							
Years	0.22	0.23	0.06	0.06	0.24	-0.09	0.00	0.07
Std. Err	0.21	0.22	0.24	0.32	0.22	0.23	0.16	0.18
N	1021	1021	1040	1040	729	729	885	885
p-value of diff.	0.63	0.50	0.12	0.12	0.94	0.30	0.62	0.27
Bandwidth	30	30	30	30	21	21	26	26

Notes: $Panel(a)$ reports results from two FRD regressions assessing the effect of an additional year of schooling on each cognitive outcome. RD1 adjusts for the following covariates: a linear trend in month-year of birth cohort (interacted with the reform dummy), indicators for gender, interview month, interview year and a quadratic term in age. RD2 additionally adjusts for month-of-birth dummies. The standard errors are clustered by month-year of birth cohort. Survey weights adjusting for unit non-response and sample design were used in all specifications. $Panel(b)$ presents the first-stage statistics. First-stage $\hat{\beta}$ is the coefficient on the reform dummy in the first stage regression. F-statistic is the robust F-statistic for the first-stage. $Panel(c)$ shows the sample size. $Bandwidth$ refers to the number of month-year of birth cohorts included in the estimation sample on each side of the treatment cut-off, selected using a data-driven procedure. N is the sample size used in each regression.

Figure 1.8: Sensitivity analyses



Notes: The bold black line plots coefficients from a fuzzy Regression Discontinuity Design model of the effect of years of school on each cognitive outcome (Word Recall, Verbal Fluency, Numeric Ability, Serial Subtraction) across a range of bandwidth choices (12 month-year of birth cohorts to 60 month-year of birth cohorts). The sample window is determined by bandwidth, measured in month-of-birth cohorts included either side of the 01 September 1957 treatment cut-off. The grey areas depict 90%, 95% and 99% confidence intervals around the treatment effect. The vertical line indicates the optimally chosen bandwidth.

Table 1.7 reports the main results computed for the full sample now using a quadratic specification in the running variable. This specification is less reliable than the linear case: the optimal bandwidth is larger, and comparison is being made between persons born too far apart to be considered comparable. Additionally, recent work have made the case that higher order polynomials are a unreliable specification in RD applications (Gelman & Zelizer, 2015) and generally favour the local linear approach (Cattaneo *et al.* , 2017b).

Table 1.7: OLS and RD estimates of the effect of schooling on cognitive outcomes (full sample, i.e., left school $\leq$ 19 years, quadratic running variable)

	Outcome							
	Word recall		Verbal Fluency		Numeric Ability		Subtraction	
	RD1	RD2	RD1	RD2	RD1	RD2	RD1	RD2
(a) RD treatment effect								
Years	0.44	0.61	0.41	0.78	0.10	0.68	0.06	-0.09
Std. Err	0.17	0.22	0.35	0.52	0.16	1.16	0.17	0.44
F-statistic	21.6	18.0	9.0	4.4	3.8	0.3	5.8	1.1
Bandwidth	52	52	34	34	27	27	30	30
N	4044	4044	2667	2667	2096	2096	2307	2307
(b) RD treatment effect (Women)								
Years	0.65	0.90	0.60	1.67	0.61	5.29	-0.08	4.67
Std. Err	0.29	0.39	0.61	2.02	0.34	37.02	0.72	32.63
F-statistic	6.6	4.9	2.7	0.8	1.7	0.0	0.8	0.0
Bandwidth	52	52	34	34	27	27	30	30
N	2267	2267	1489	1489	1166	1166	1286	1286
(c) RD treatment effect (Men)								
Years	0.35	0.43	0.27	0.42	0.61	-0.43	0.10	0.23
Std. Err	0.24	0.29	0.39	0.60	0.34	1.13	0.23	0.38
F-statistic	17.2	13.3	6.3	4.9	1.7	0.4	5.1	2.1
Bandwidth	52	52	34	34	27	27	30	30
N	1777	1777	1178	1178	1166	930	1021	1021
p-value of diff.	0.63	0.50	0.12	0.12	0.94	0.30	0.62	0.27
Bandwidth	30	30	30	30	21	21	26	26

Notes: $Panel(a)$ reports results from two FRD regressions assessing the effect of an additional year of schooling on each cognitive outcome. RD1 adjusts for the following covariates: a quadratic trend in month-year of birth cohort (interacted with the reform dummy), indicators for gender, interview month, interview year and a quadratic term in age. RD2 additionally adjusts for month-of-birth dummies. The standard errors are clustered by month-year of birth cohort. Survey weights adjusting for unit non-response and sample design were used in all specifications. $Panel(b)$ presents the first-stage statistics. First-stage $\hat{\beta}$ is the coefficient on the reform dummy in the first stage regression. F-statistic is the robust F-statistic for the first-stage. $Panel(c)$ shows the sample size. $Bandwidth$ refers to the number of month-year of birth cohorts included in the estimation sample on each side of the treatment cut-off, selected using a data-driven procedure. N is the sample size used in each regression.

Appendix 1.D Schooling's effects on occupation

Table 1.8 examines the effect of an additional year of schooling on the intermediate outcomes. It shows the results of three models: the causal effect of an additional years of schooling on occupation type of the first job after leaving school, occupation type of the current job, and technical skill usage. The first two specifications, which have categorical outcomes, are estimated via Maximum Likelihood using the *cmp* package in Stata. The third specification, with a continuous outcome, was estimated using 2SLS. The first column in Panel (a) shows the effect of taking 16, rather than 15 years of schooling on the average probability of choosing each occupation type in one's first job. An additional year of schooling leads to statistically significant reduction in the probability of taking a routine or semi-routine occupation immediately after leaving school, of 11 percentage points. On the other end of the spectrum, an additional year of schooling leads to an average of a 4 percentage point increase in the probability of taking a managerial or professional job immediately after leaving school.

As shown in Panel (b), the effects on current occupation (measured toward the end of working life) are smaller in magnitude: for instance, an additional year of schooling leads to 16 percentage point reduction in the average probability being in a routine or semi-routine occupation. Panel (c) presents the results of a 2SLS regression examining the effect of schooling on the extent of technical skills used in first job after leaving school. An additional year of school increases the index of technical skills used in first job out of school by 1.72 units, the equivalent of half of a standard deviation.

Table 1.8: Causal effects of one additional year of schooling on intermediate variables

Outcomes	Average marginal effects	Standard error
	Panel (a)	
NS-SEC (first job)		
Routine; semi-routine	-0.11	0.02
Lower supervisory and technical	-0.01	0.00
Small employers and own account	0.01	0.00
Intermediate occupations	0.07	0.01
Managerial and professional	0.04	0.00
	Panel (b)	
NS-SEC (current job)		
Routine; semi-routine	-0.16	0.01
Lower supervisory and technical	-0.01	0.00
Small employers and own account	0.01	0.00
Intermediate occupations	0.02	0.00
Managerial and professional	0.14	0.01
	Panel (c)	
Technical skills		
Cognitive skills	1.72	0.80

Notes: This table reports results from three separate instrumental variable (IV) procedures, assessing the effect of an additional year of schooling on the intermediate outcomes using the RoSLA as the IV. $Panel(a)$ shows results from an IV ordered probit model, with occupation type of first job as the outcome. $Panel(b)$ shows results from an IV ordered probit model, with occupation type of current job as the outcome. These two specifications were estimated using Maximum Likelihood via the $-cmp-$ Stata package. The third specification, in $Panel(c)$ shows results from a 2SLS regression with Technical Skill usage as the outcome. The covariates in each specification are: a linear trend in month-year of birth cohort (interacted with the reform dummy), indicators for gender, interview month, interview year, and a quadratic term in age. The standard errors are clustered by month-year of birth cohort (the running variable). Survey weights adjusting for unit non-response and sample design are used in all specifications.

Appendix 1.E Estimation of direct and indirect treatment effects

The following expressions describe the estimators for the Local Average Complier Mediated Effect (LACME), $\delta(d)$, and the Local Average Natural Direct Effect (LANDE), $\zeta(d)$, developed in Yamamoto (2014); Park & Kürüm (2018). For the categorical mediators (occupation type), the integral is replaced by a summation.

$$\tilde{\delta}(d) = \frac{1}{N}\sum_{i=1}^{N}\int \frac{\tilde{Q}_{d|dX_i}\tilde{G}_{m|ddX_i}S_{mddX_i} - \tilde{Q}_{d|(1-d)X_i}\tilde{G}_{m|d(1-d)X_i}\tilde{S}_{md(1-d)X_i}}{\tilde{Q}_{d|dX_i}\tilde{G}_{m|ddX_i} - \tilde{Q}_{d|(1-d)X_i}\tilde{G}_{m|d(1-d)X_i}} \times \frac{(\tilde{Q}_{1|1X_i}\tilde{G}_{m|11X_i} - \tilde{Q}_{1|0X_i}\tilde{G}_{m|10X_i} - \tilde{Q}_{0|0X_i}\tilde{G}_{m|00X_i} + \tilde{Q}_{0|1X_i}\tilde{G}_{m|01X_i})}{\tilde{Q}_{1|1X_i} - \tilde{Q}_{1|0X_i}}\, dm, \quad (1.13)$$

$$\tilde{\zeta}(d) = \frac{1}{N}\sum_{i=1}^{N}\int \left\{\frac{\tilde{Q}_{1|1X_i}\tilde{G}_{m|11X_i}S_{m11X_i} - \tilde{Q}_{1|0X_i}\tilde{G}_{m|10X_i}\tilde{S}_{m10X_i}}{\tilde{Q}_{1|1X_i}\tilde{G}_{m|11X_i} - \tilde{G}_{m|10X_i}} - \frac{\tilde{Q}_{0|0X_i}\tilde{G}_{m|00X_i}\tilde{S}_{m00X_i} - \tilde{Q}_{0|1X_i}\tilde{G}_{m|01X_i}\tilde{S}_{m01X_i}}{\tilde{Q}_{0|0X_i}\tilde{G}_{m|00X_i} - \tilde{Q}_{0|1X_i}\tilde{G}_{m|01X_i}}\right\} \times \frac{(\tilde{Q}_{d|dX_i}\tilde{G}_{m|ddX_i} - \tilde{Q}_{d|(1-d)X_i}\tilde{G}_{m|d(1-d)X_i}}{\tilde{Q}_{1|1X_i} - \tilde{Q}_{1|0X_i}}\, dm. \quad (1.14)$$

where, $\tilde{S}_{mdzx} = E[Y_i \mid M_i = m, D_i = d, Z_i = z, X_i = x]$, $\tilde{G}_{m|tzx} = p(M_i = m\,|D_i = d, Z_i = z, X_i = x)$, $\tilde{Q}_{d|zx} = Pr(D_i = d\,|Z_i = z, X_i = x)$.

A linear model, estimated using least-squares, was employed for the treatment, the continous mediator (intensity of technical skills used in first job) and the outcome (Word Recall). The covariates in the treatment equation were the instrument and pre-treatment covariates; the covariates in the mediator equation were the treatment, instrument, an interaction between treatment and instrument, and pre-treatment covariates; in the outcome equation the covariates were main effects and interactions between the mediator, treatment and instrument, and pre-treatment covariates.

An ordered probit model was employed for categorical mediators (occupation type), conditioning on treatment, instrument, an interaction between treatment and instrument, and the pre-treatment covariates. The pre-treatment covariates included were: a linear trend in month-year of birth cohort (interacted with the reform dummy), indicators for gender, interview month, interview year, and age-at-interview.

Appendix 1.F Causal mediation analysis: sensitivity to additional covariates

Table 1.9: Causal mechanisms: LACME and LANDE (highest parental occupation (SOC2000))

	Occupation (first job)		Occupation (current job)		Technical skills	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
LATE	0.53	(0.39)	0.53	(0.42)	0.53	(0.41)
LANDE						
$\tilde{\zeta}(0)$	0.45	(0.82)	0.40	(0.86)	0.53	(0.87)
$\tilde{\zeta}(1)$	0.53	(0.55)	0.54	(0.48)	0.43	(0.49)
LACME						
$\tilde{\delta}(0)$	0.00	(0.40)	0.04	(0.34)	0.00	(0.41)
$\tilde{\delta}(1)$	0.08	(0.74)	0.04	(0.79)	0.10	(0.75)
Bandwidth	30	30	30	30	30	30
N	939	939	939	939	939	939

Notes: This table shows estimates of the local average treatment effect (LATE) of one additional year of schooling on Word Recall, and decomposes this into components explained and unexplained by three alternative intermediate variable (Occupation type of first job, Occupation type of current job, Technical skills in first job). $\tilde{\zeta}(0)$ and $\tilde{\zeta}(1)$ ($\tilde{\delta}(0)$ and $\tilde{\delta}(1)$) denote estimates of the LACME (LANDE) when the treatment is fixed to either $d = 0$ or $d = 1$, respectively. Each model adjusts for the following covariates: linear trend in month-year of birth cohort interacted with the reform dummy, indicators for gender, interview month, interview year, age. $Bandwidth$ refers to the number of month-year of birth cohorts included in the estimation sample on each side of the treatment cut-off.

Appendix 1.G Causal mediation analysis: sensitivity analyses to unmeasured confounding

This section explores the sensitivity of the findings to mediator-outcome confounding. VanderWeele (2010, 2015) develop a formula to acsertain the effects of a hypothesis level of confounding on the mediated effect. This approach has been applied in various applications, including Ananth & VanderWeele (2011), and in Park & Kürüm (2018) (who develop the estimation approach used in this paper).

In VanderWeele (2010, 2015)'s method, the impact of mediator-outcome confounding on the estimate of the indirect effect, $\delta(t)$, can be split into two pathways. First, denoted β, is the association between the instrument (and hence the treatment, given an excludability assumption) and the unobserved covariates via the mediator. Second, denoted α, the association between the unobserved covariate and the outcome. Assume these associations do not vary by the value of the observed covariates (although this could, in theory, be relaxed). The biased estimate of $\delta(t)$ is expressed as $\delta(\hat{t}) = \delta(t) + \alpha\beta$. Re-arranging gives the true estimates for the confounding effects of the unobserved covariate U, the estimate of the indirect is $\delta(t) = \delta(\hat{t}) - \alpha\beta$

A useful exercise is to consider how large the effects of the unobserved covariate would need to be, such that the true estimate of $\delta(t)$ would be zero. As an exemplar, I take the estimate $\delta(\hat{1})$=10%—i.e., how much of the total effect of education on cognitive function can be explained by working in a job relying on technical/STEM-type skills.

Considering the the first path, supposing that a person with a high value of the unobserved covariate has high working memory by 0.15 standard deviations (selected based on the largest associated between any observed covariate the outcomes, conditioning on pre-treatment covariates and occupation choice—which is from the lagged value of the outcome), for any given values of their occupation and pre-treatment covariates. Suppose also that persons who stay for an additional year of schooling had a higher probability of having a high value of the unobserved covariate of 60% (compared to those who leave at 15 years), then this level of confounding would mean the unbiased estimate would be zero

(0 = 0.10 - (0.6)(0.15)). This level of confounding could be considered very high, in the economic and social sciences, and likely implausibly in this scenario. A more moderate level of confounding which would reduce but not completely remove the effect size of 10% is more plausible.

Chapter 2

The long run effects of missing out on a preferred secondary school

co-authored with Ian Walker

The authors are grateful for many helpful discussions with Matthew Weldon, and comments from participants at 3rd IZA Workshop: The Economics of Education 2018 and the VUW Applied Econometrics Workshop 2018, and acknowledge funding from the Economic and Social Research Council. Thanks go to John Coldron and colleagues for providing the data from their survey of school admission policies. Thanks to the UK Data Service for their provision of the secure data access and support for datasets SN7104 and SN8189.

2.1 Introduction

Demand for places at popular schools exceeds supply, and not all families are able to attend their preferred school—yet attending a preferred school may have potentially important consequences for long term outcomes, if parents choose schools based on test score improvements or other school attributes important for child outcomes. In addition to academic outcomes, parents may rationally value wider factors provided by schools, for example non-cognitive skill development, peer groups and pupil well-being, which can have important consequences for success in life (Heckman *et al.* , 2006; Mendolia *et al.* , 2018).

Evidence to date has revealed mixed findings on whether parents choose schools based on a causal improvements in test scores, rather than peer mix, pedagogy or other school attributes (Abdulkadiroglu *et al.* , 2017a). Consequently the effects of attending a preferred school are also mixed, with the majority of studies documenting little effect of attending a preferred school on short-run test scores (Cullen *et al.* , 2006; Dobbie & Fryer, 2014; Deming *et al.* , 2014; Abdulkadiroglu *et al.* , 2014; Hoekstra *et al.* , 2018; Abdulkadiroglu *et al.* , 2017b, 2018). These studies are typically based on data from specific cities (for example, Boston or New York), and specific settings, such as attendance at academically "elite" schools.

In England, who gets into their preferred secondary school is a well-publicised issue in the media. We know that parents value secondary schools which perform well on common academic metrics (Burgess *et al.* , 2017; Gibbons & Silva, 2011), and there is regional and demographic patterning in the share of families gaining entrance to their preferred school (Weldon, 2018). Yet to date there is no evidence on the causal effects of attending (or not) their preferred school. This is in spite of the fact that the UK provides an ideal laboratory to explore the effects of schooling, due to a centrally managed quasi-market for secondary schools and high quality linked administrative data. Our paper contributes new evidence by combining powerful administrative data and detailed cohort data, which follows pupils from secondary school to age 25 years, to study the long-term consequences of attending a preferred school.

Generally speaking, how pupils are matched to schools is governed by the so-called "school choice" problem. This refers to the mechanism that matches pupils and schools, given parental preferences, school capacities and other policy priorities—for instance, diversity in schools or social mobility. A mature literature on the market design approach to school choice has developed over the last decades, which considers how to best elicit parental preferences over schools and then assign places to children (Abdulkadiroglu & Sönmez, 2003; Cantillon, 2017).

The way this system operates in England is that families submit their ranked preferences over schools to the local government authority, each school ranks families based on a set observable characteristics (such as distance from the school), and then this information is combined to offer places parents, subject to school capacity constraints. High-performing secondary schools tend to be oversubscribed: there are more applicants than places available. In the case of oversubscribed schools, places are rationed based on a set of published criteria, such as whether the applicant has older siblings at the school, and the family's distance from the school (West *et al.* , 2011).

A number of empirical regularities have been established in England. We know that, on average, parents value schools perceived as academically "good" schools, and are willing to travel to attend them (Burgess, 2015; Burgess *et al.* , 2017). There are demographic and socio-economic differences in engagement with the school choice process. For example, families of minority ethnic groups engage more with the school choice process than White families: they are willing to travel further to attend a high quality school, and they list more schools in their submission of preferences.

However, less affluent pupils, and ethnic minority groups, are less likely to attend good schools than their peers, and have poorer performance in high stakes exams, and subsequently entry into elite higher education. This motivates current UK government priority is to broaden access to high performing schools, as recently recognized in the Government's *Unlocking Talent, Fulfilling Potential: A plan for improving social mobility through education*, "In Britain today, the community where you grow up will shape your chances of attending a good school and your wider educational and career outcomes."(Department for Education, 2017). School admissions may represent one component of the complex reasons for variation in attainment in England, as the socio-economic and ethnic variation in who attends good schools is not fully explained by variation in preferences and residential location (Weldon, 2018).

While many changes in the operation of the school choice system have occurred over recent decades (West *et al.* , 2011), the share of children being offered a place at their first choice school has remained static at about 84%. Taking the stock of "good schools" as fixed, this statistic has both equity and allocative efficiency concerns at its core. First, if particular segments of society more likely to miss out on a place at a preferred school, this raises and equity concern about equal access to good schools. In England, there are large differences in attending a good school by ethnic group, which is unlikely to be explained by differing preferences. A second possibility is school-pupil match effects, which would mean the productivity of school varies depending on its intake, and matching the right pupils to the right schools improves school performance and child outcomes.

Outside the UK, there are a number of studies combining market design theory with modern methods for policy evaluation to assess how missing out on a preferred school impacts outcomes (Pop-Eleches & Urquiola, 2013; Cullen *et al.* , 2006; Dobbie & Fryer, 2014, 2015; Abdulkadiroglu *et al.* , 2017b), often with a focus on peer effects as a possible channel (Abdulkadiroglu *et al.* , 2014; Hoekstra *et al.* , 2018). Peer effects are an appealing channel through which preferred schools could exert and effect on outcomes, especially if they are dynamic in nature. Somewhat counter-intuitively, the majority of these studies do not find any causal effect of attending a preferred secondary school in academic outcomes. This puzzle could have a number of possible explanations. Many of these studies use a compelling empirical design, comparing the outcomes of pupils who "just" got into a school compared with those who just missed out—either side of a threshold based on a continuous tests score.

There are at least two possible offsetting effects. First, pupils who just get will be at the bottom of the ability rank in a top tier school, whereas those who miss out will be at the top of the ability rank. Rank has been documented to have a direct effect on outcomes (Elsner & Isphording, 2017). Second, parents and families may increase (reduce) their effort invested in the student if they get in (misses out), again exerting an offsetting effect (Pop-Eleches & Urquiola, 2013). More generally, parents may not choose schools based on school effectiveness; either because they are not aware of the effectiveness of different schools, or they mistake the quality of the intake mix for school effectiveness (Abdulkadiroglu *et al.* , 2017a).

On the other hand, parents may value other school attributes, such as "elite" status of the school, diversity in the school, or expected skill improvements on other margins, such as non-cognitive skills. Indeed, choosing a school based on considerations other than test scores can certainly be a rational choice, which would show up in a wider set of outcome, such as well-being and mental health, and longer-term outcomes. For instance, Jackson & Beuermann (2018) provide compelling evidence that while attending preferred school in Barbados does not improve test scores, it does reduce early pregnancy and enhance employment rates among women.

We contribute by providing the first evidence to date on on the long-run consequences of attending a preferred school in England. England is a useful laboratory to explore this topic for a number of reasons. First, nationwide data are available and school finances are determined centrally, allowing us to study this topic at the national level. Second, given parental preferences and school capacities, two different mechanisms for allocating school places were deployed across different areas at the time that the children in our cohort were making their school selections. This allows us to estimate the treatment effects separately for two mechanisms which have been studied widely in the theoretical literature (Pathak & Sönmez, 2013). Third, for most schools in England selection on ability is explicitly prohibited—meaning we escape the confounding effect of ability rank faced by other studies, and our estimates are more generalisable across the ability distribution.

Another advantage is that our data capture long term outcomes observed at age 25 years as well as the typical short run educational outcomes which, in our case, are high-stakes in that they play a decisive role in dropping out from school (i.e. leaving at the minimum school leaving age because one is unable to gain admission to post compulsory schooling) and in facilitating access to higher education.

We employ the Longitudinal Study of Young People in England (LSYPE, also known as *Next Steps*), a nationally representative birth cohort study which tracks the lives of a cohort of around 15,000 young people in England who were born in 1989/90 (Henderson, 2017). LSYPE began when the children were in Year 9, the second year of secondary school, and follows them until they are aged 25 years (annually until age 19/20, then a gap until age 25/26 years).

The LSYPE contains detailed information on school choice, including whether the child attends their preferred school, family background, experiences in school and, crucially, labour market and university outcomes. This data is confidentially linked to the National Pupil Database (NPD), a database of administrative school records, containing individual test scores and child characteristics, as well as school-level attributes. From these data we can characterize the attributes of schools, preferred and otherwise, and track the achievement of the LSYPE pupils in high-stakes examinations (the GCSE, taken at age 15/16, and A-levels taken at age 17/18).

To apply for a place at a state school in England (about 90% of schools during our time period), parents submit a ranking of their preferred schools. Parents can list between 3 and 6 places depending on their local area. For oversubscribed schools, allocation of places is prioritised based on a set of observable criteria, typically including: whether the child is categorised as "looked after"; whether the child has special education needs; whether the child has an older sibling at the school; and, finally, the distance from the school. In some cases, faith may also be considered. If a pupil cannot be allocated to any of their listed preferred school they are assigned to a school with spare capacity (which is typically of lower quality). For example, two families who look observationally similar may have different rates of acceptance due to small—arguably random—differences in their proximity to their preferred school.

Based on the literature on parental choice and the details of the institutional setting, we build a case that the child's potential outcomes are independent of whether they get into their first-choice school after conditioning on variables determining school admission, proxies for parental preferences, detailed information on socio-economic status, prior ability of the child and the feasible choice set of schools. After adjusting for these factors, we argue that missing out on a preferred school is conditionally independent of the child's outcomes.

In terms of estimation, we use a combination of matching and regression adjustment to estimate the average effect of attending a preferred school. First, we use kernel matching to create a matched sample of pupils, matching each pupil who missed out on a place at their preferred school, to a set of control pupils who got in—and who live in close proximity to the treated pupil's school. The matches are selected based on their straight line distance from the treated pupil's school, within an optimally selected radius. Using this geographically matched sample, we use regression-adjustment to estimate the average treatment effect of attending a preferred school, where the contributions of the control observations are weighted in inverse proportion to their distance from the treated unit's school. The purpose of pre-processing the data via the matching is to balance the sample, such that both the control and treated units have a similar choice set of schools. This reduces reliance on the functional form of the regression to capture geographical factors.

Average treatment effects are likely to hide variation in treatment effects, as the consequences of attending a preferred school are likely to differ across the population. We explore variation in treatment effects by gender and assignment mechanism. When our cohort were applying for schools, two mechanisms to allocate places (given preferences and school capacities) were used by admissions authorities: so-called Deferred Acceptance (DA) and First Preference First (FPF) approaches. The mechanism deployed differs by local area. This allows us to estimate the treatment effects for these two different algorithm designs that have been most studied in the mechanism design literature (Pathak & Sönmez, 2013).

We unpack the effects of attending a preferred school by investigating a number of potential mechanisms. The average effect of missing out on a preferred school is informed by comparisons both within and between schools. Pupils who miss out may have poorer outcomes due to attending a lower quality school, and/or due to a direct effect of "not attending a preferred school"— a "match effect". We control for school fixed effects to assess the extent to which the effects of attending a preferred school is due to between-school variation in school attributes, and within-school variation—potential match effects. Second, we considering a number of intermediate outcomes—such as mental well being, and engagement in risky behaviours.

Pupils who miss out on a preferred school are more likely to be a member of ethnic minority groups, are less likely to have older siblings (a key determinant of admissions), and there is negligible difference in prior ability. Preferred schools have higher headline measures of attainment, and tend to be more affluent, on average. Our analysis fails to detect convincing evidence for effects of missing out a preferred school on test scores and university attendance. However, we do find consistent evidence of increased high school drop-out after the age of compulsory schooling, and increased university drop-out among pupils who miss out on a preferred school. These negative effects persist to age 25 years, where we find long-lasting negative effects on mental health and income.

Pupils who miss out show increased patterns of early engagement in risky behaviours, representing a potential mechanism driving the increased drop-out rates. While we do not find strong evidence for gender differences, the negative consequences of missing out are more pronounced in local areas which used a more manipulable mechanism ('first-preference first') for allocating school places. Pupils exposed to the first-preference mechanism who miss out end up commuting further to school than their counterparts exposed to a Deferred Acceptance mechanism, possibly indicating much poorer school matches. Overall, our findings highlight that which school a child attends can have consequences for broader outcomes than only test scores, and the way places are allocated matters for ensuring children attend a school which is right for them.

2.2 Institutional setting

2.2.1 School types and curriculum stages

Ages and stages:

The English school curriculum is divided into blocks of years called "key stages" (KS). At the end of each KS, performance is examined via examination. The final three years of primary school form Key Stage 2, which is assessed in a test at age 10/11 years. Children now transition to secondary school, where they must attend until age 16 years. Key Stage 3 runs from Year 7 to Year 9. Key stage 4 follows from year 10 and 11, culminating in the high stakes General Certificate of Secondary Education examinations (GCSEs) exams. Pupils typically take between 5 and 10 subjects.

At this point, with satisfactory performance in GCSEs, pupils can study for Advanced levels (A levels), from age 16 to 18 years. A-Levels are the most common route to university entrance, and are taken either at the same secondary school, or some pupils move to another school or to a specialised "sixth form college" to take their A levels. Further education (FE) colleges provide vocational training, often pursued in combination with employment or A-levels. Pupils typically enter Higher Education (University) from age 18. Higher Education participation rates are over 40% of the cohort and this has grown dramatically in the last two decades.

School types:

There were different types of secondary schools in England in 2001, described as follows. The first type are "state schools", which comprise about 92% of schools, and are controlled and funded by the government. All children in England between the ages of 5 and 16 are entitled to a free place at a state school. These schools all follow the National Curriculum and are inspected by Ofsted (the government's Office for Standards in Education, Children's Services and Skills). The second type is "Independent Schools". Around 7% of families choose to pay for a place at an Independent (also called a private or public) school. Our analysis is based on pupils who attend state schools (starting school in September 2001), rather than independent schools.

Within the category of state schools, there are community schools (69%), voluntary-controlled (3%), voluntary-aided (14%) and foundation schools (14%) and Grammar schools. *Community schools* are entirely run by the local council (Local Authority). *Foundation* *or Trust schools* are run by a local governing body. These schools were formerly called "Grant-maintained" schools–this was an initiative to allow more freedom in provision of education where, by majority parental vote, schools could opt-out of Local Authority control and be run by a governing body with more control over admissions and staffing. *Voluntary-aided schools* are typically religious or faith schools, which can admit pupils on religious affiliation grounds. *Voluntary-controlled schools* are almost all faith schools. They are a mix between community and voluntary-aided schools: similar to a community school, the local authority employs the staff and sets the entrance criteria, but the school land and buildings are owned by a charity, often a church, which also can appoint some members of the governing body. Regardless of governance arrangement, all state schools have to comply with the school admissions code of practice which sets guidelines for fair admissions. Finally, there are some remaining *Grammar schools*, again run by the council, a foundation body or a trust - they select all or most of their pupils based on academic ability based on the so-called 11 plus exam.

2.3 School choice in England

"School choice" is defined as policies which allow parental preferences to influence which school their child will attend. The central problem in designing a school choice system is in allocating pupils to schools, accounting for both parental preferences, school capacities and wider policy objectives (e.g. diversity in schools). One motivation for school choice focuses on the importance of respecting parental preferences, and allowing them to choose from a diversity of school types. A second motivation, not necessarily implied by the first, is to induce competition among schools with the aim of improving school standards. Both of these motivations have been referenced by different UK Governments in relation to school choice policy.

The rationale behind using school choice to enhance competition, is that by allowing families to choose which school they attend, popular schools will flourish, and poor performers will be forced out of this "quasi-market". A number of conditions are required for this system to work, including: (i) parents engage with this process, they can observe the quality of different schools, and they choose schools based on effectiveness; (ii) there is a choice of schools available to parents; (iii) good schools can expand to meet increased demand; (iv) and poor performing schools face consequences (they are shut down or provided with remedial assistance).

In England, features of parental "choice" and a "quasi-market" in compulsory schooling were introduced largely from 1988, formalising the role of parental preferences, awarding extra funding for popular schools, and allowing school some discretion in the pupils they admitted. The specific operation and political motivation of the school choice setting in England has evolved substantially over the last few decades. Very generally, the process to allocate places works as follows: to apply for a place at a state school in England, parents submit a ranking of their preferred schools to the local government. They can list between 3 and 6 places depending on their local area.

For oversubscribed schools, allocation of places is prioritised based on a set of observable criteria ("school priorities"), typically including: whether the child has exception "medical or social needs"; whether the child has an older sibling at the school; and, finally, the distance from the school as the tie-breaker. In some cases, faith may also be considered. An algorithm is used to allocate school places, given parental preferences, school priorities and school capacities (for example, a Deferred Acceptance algorithm is used). If a pupil cannot be allocated to any of their listed preferred schools (often because the parents do not list enough schools), they are assigned to a school with spare capacity (which is typically a less desirable school).

The LSYPE cohort were applying to schools in 2000, to start in September 2001, when the admissions arrangements were less regulated, and subject to less stringent legislation, compared with today. The following section describes in detail the policy settings applicable for the LSYPE cohort.

Admissions and oversubscription criteria used in September 2000

The admission process varied both within and between local authorities, and in September 2001 was more heterogeneous and under-developed compared with the present day arrangements. At its most general level, the admissions process had three stages (i) parents submit their preference(s) to the admission authority (or authorities), (ii) the admission authority tries to assign families to their preferred schools, and (iii) in the case of over-subscription, applications are ranked based on a set of observable criteria to ration scarce places. The School Code in force at the time provided some suggestions for the specific criteria to be used: *"Commonly used and acceptable criteria include sibling links, distance from the school, ease of access by public transport, medical or social grounds, catchment areas and transfer from named feeder primary schools, as well as parents' ranking of preference"* (DfE (2001), para. 3.14). In practice, the actual procedures followed vary and are difficult to categorise (see West *et al.* (2004); Williams *et al.* (2001) for surveys of the admission protocol used, and specific examples from LEA brochures).

Once the families' preference rankings have been submitted to an admissions authority, a procedure is required to allocate school places, given school capacities and priorities. There are a number of approaches for doing this. The two used in 2000 were: the FPF *first-preference* algorithm, also called the Boston or immediate-acceptance mechanism; and the *equal-preference* algorithm, also called the Gale-Shapley, or deferred-acceptance (DA), mechanism.

The FPF approach proceeds in rounds. First, each school starts by assigning places to pupils *who put that school as their first choice*, based on the ranking determined by the school's admission criteria—until either there are no places left, or all students who put that school as first choice are assigned. The next round conducts the same procedure for pupils who put that school as their second-choice, and so on. This approach maximises the share of families getting their first-choice school. The drawback of this approach is that families who are not offered a place at their first preference school can also be rejected by their second preference school in favour of a child of parents who placed that school first, despite living further away. If families put their true preferred school as their first choice, but where they have a very low chance of acceptance, they may not only fail to secure a place at their preferred school, but also fail to be placed at any school they deem acceptable. This means that families have an incentive to trade-off their true preferences against their probability of acceptance, in order to increase the chance of gaining a place at a satisfactory, but not most preferred, school. 41% of schools used a 'first-preference' approach to offer places for the September 2001 intake (West *et al.* , 2004).

In contrast, the DA approach does not use the preference ranking as a method for schools to rank applicants. In this method, parents submit their preference ranking over schools. The families' application is considered at each school that they listed as a preference without reference to its ranking. Places are offered to applicants who fulfil the oversubscription criteria of a school to the greatest extent. This may lead to a parent whose child does not meet the criteria of his first preference school being offered a place at a nearby school for which he has named a second preference. Where proximity is a key criterion, one effect of this system is to confer advantage to parents who live near to several popular schools.

West *et al.* (2004) used survey evidence to explore the use of various admission criteria and identified older siblings at the school (96%), distance from home to the school (86%), medical/social need (73%), catchment area (61%) and 'first preference' (41%) as the most common admission criteria practices (see Appendix 2.E, Table 2.20 for a longer list). In 2000, it was not required to give "looked-after" children first priority. From 2006, it was a statutory requirement that children in care should be given top priority in the event of a school being oversubscribed, and in 2008, almost all schools (99%) had an admissions criterion relating to children in care compared with 2% in 2001 (West *et al.* , 2011).

As described further in Appendix 2.E, there are a range of sources of uncertainty about whether one will be admitted to a particular school. Catchment area policies, while becoming less popular coming up to 2001, were used in both general admissions (14.9%) and featured somewhere as an oversubscription criteria in (63.1%) of LEAs Williams *et al.* (2001). Catchment area policies create uncertainty because the catchment areas can change over time in unpredictable ways. For example, in the survey by Williams *et al.* (2001), one LEA deployed a flexible catchment area system based on the numbers of children applying to particular secondary schools from particular primary schools over a period of time. Additionally, sibling and catchment criteria can interact; for example, siblings in a catchment area, other children living in a catchment area, siblings living outside a catchment area and, finally other children outside a catchment area. Given random fluctuations in the sibling distribution, this creates some random variations in the effects of the catchment area on the probability of admission.

The changing interaction of sibling and distance criteria makes it difficult for parents to predict their chances of gaining a place at a particular school, complicated further by the interaction with banding by ability. Distance to school is itself an unpredictable criterion, because families do now know in advance whether or not they live close enough to a school in any given year, because the distance cut-off changes every year due to population fluctuations, changes in siblings at the schools, house-building programmes, changes in LEA boundaries, and school closures. Although some individual schools, in some LEAs, publish information over several years about the 'cut-off point' for distance, typically it is not clear how near 'near enough' will be the next year (Williams *et al.* , 2001). The least predictable factor is the FPF system because it depends, among other things, upon how many parents apply to a school in a given year, which fluctuates year-on-year and cannot be known in advance (Williams *et al.* , 2001).

A final important issue is the interaction between the non-selective public school sector with grammar schools and private schools. Appendix 2.E provides detail about the various ways in which the Local Authority preference lists interacted with the grammar school admission process, and shows that these arrangements varied by area. Sensitivity analyses to how these outside options may impact the analyses in this paper is a area which will be pursued in further extensions to this work.

2.4 Data

2.4.1 Longitudinal Study of Young People in England

We employ the Longitudinal Study of Young People in England (LSYPE, or *Next Steps*), a nationally representative birth cohort study which tracks the lives of a cohort of around 15,000 young people in England who were born in school year 1989/90 (Department for Education, 2011; Henderson, 2017; University College London, 2018). The study begins when the children are in Year 9, the third year of secondary school, and follows them until they are aged 25 years (annually until age 19/20). The LSYPE contains detailed information on school choice, including whether the child attends their preferred school, as well as family background, experiences in school, and crucially labour market and university outcomes.

These data are confidentially linked to the National Pupil Database (NPD) (linkage rate = 97%), a database of administrative school records. The NPD contains individual test scores, individual characteristics, and school-level attributes (e.g., socio-economic and ethnic mix). From this data we can track the achievement of the LSYPE pupils in statutory examinations: primary school tests taken at age 10/11 (Key Stage 2), tests taken at age 13/14 in secondary school (Key Stage 3); the high stakes national examinations taken at age 15/16 in at least five subjects, including English and Mathematics (General Certificate of Secondary Education (GCSE); finally, at age 17/18, further national examinations known as A(dvanced)-levels are taken usually in three subjects.

The LSYPE sampling frame comprised all pupils attending maintained schools, independent schools, and pupil referral units (attended by the most challenging students) in England in February 2004. The first 7 waves of the study were funded by the Department for Education (DfE), commissioned to explore the factors shaping educational attainment and transitions out of compulsory schooling. The final wave, at age 25 years, was funded by the Economics and Social Research Council.

The LSYPE used a two-stage sampling design, first sampling schools, then pupils within schools. Deprived schools, defined as those in the top quintile of the free school meal receipt distribution, were oversampled by a factor of 1.5. Pupil selection probabilities were dependent on ethnic group (White; Indian; Pakistani; Bangladeshi; Black African; Black Caribbean; and Mixed), and on school selection probabilities, aiming to sample at least 1,000 pupils in each ethnic group. The interviews in Waves 1 to 4 were carried out face-to-face at the young person's home, using computer assisted personal Interview (CAPI) software, and interviewed both the young person and the main parent. From Wave 5 onward, the survey used mixed-methods (online, telephone and face-to-face interviews).

Table 2.12 shows the structure of the LSYPE, the sample size in each wave, response levels and fieldwork timing. The fieldwork was mainly carried out between April and October, with the bulk occurring toward the end of the school year, i.e. April to August. However, a minority of the interviews fell in the subsequent academic year. The timing was designed to ensure that pupils applying to university would have passed most stages of the process and be holding offers by the time of the LSYPE interview.

In Wave 8, now managed by Centre for Longitudinal Studies, productive interviews were achieved with 7,707 cohort members, yielding a cross-sectional response rate of 51%, and the longitudinal Wave 7 to 8 figure was 69% (i.e. 69% of those responding to Wave 7 also responded to Wave 8). The LSYPE achieved cross-sectional responses rates ranging from a minimum of 51% (in wave 8), to a maximum of 92% (in Waves 3 and 4).

Weights are provided with each LSYPE wave to account for the complex survey design and survey non-response. Solon *et al.* (2015) classify the situations where weighting is appropriate into three groups: estimating population representative quantities; if missingness or survey design is correlated with the association, or causal relationship in question, and heterogeneity. In this analysis we use the weights throughout—we are interested initially in describing population quantities, such as the share of pupils who miss out on a preferred school. Additionally it seems plausible that the relationship between missing out and later outcomes could be confounded by survey design or survey drop out.

To classify local authorities by whether they used a Deferred Acceptance or First Preference mechanism, we use data collected by Coldron *et al.* (2008). This data was collected in 2006, and indicates whether a Local Authority (or in some cases specific schools) used a 'first-preference' approach. Important caveats are that, in 2000, when our cohort were applying to schools, there was less co-ordination by Local Authority in arranging admissions, therefore there was variation both within- and between- Local Authorities in use of first-preference as a mechanism—some schools were their own admission authorities. Additionally, areas which use DA in 2006 may have used FPF in 2000 and switched during the intervening years. Therefore this classification is inevitably fuzzy but still gives an indication of areas where at least some places, if not all, were likely allocated by FPF—and our subgroup estimates are in the vein of intention-to-treat estimates.

2.4.2 Outcomes

Short-run academic outcomes

- Key stage 3: Average points score from Key Stage 3 examinations
- Five A*-C: Binary variable indicating whether the pupil gained at least 5 GCSEs at grade C or above
- Stayed on: Binary variable indicating whether the pupil stayed on to study for a 'Level 3' qualification which attracts UCAS points
- UCAS points: continuous variable summing the UCAS tariff points from the pupil's best three (A-level or vocational equivalent) subjects (using the pre-2017 legacy UCAS points assignment system.)

Longer-run academic outcomes

- At uni: University attendance
- Russell group: Binary variable indicating whether the pupil attends an 'elite' Russell group university
- Drop-out (five-year completion rate: Among pupils who started university at age 18 or 19, did they complete their degree before age 25/26?
- First class: Degree class

Wider long-run outcomes

- Mental health: GHQ 12 point scales, where a higher score indicates poorer mental health
- Crime: Binary variable indicating whether the young person reports having any interaction with the criminal justice system

- Fertility: Does the young person have a child at wave 8?

- Income: Ln(annual income) at age 25 years.

Intermediate outcomes (Wave 1 and 2)

- Cannabis: ever tried cannabis?

- Any alcohol?: Drank alcohol in the past 12 months

- Truant: Has the child played truant in the past 12 months?

- Risky behaviours: sum of risky behaviours engaged in (0-12).

2.5 Empirical strategy

The key threats to credible identification of the effects of attending a preferred school can be characterised as follows:

Pupil characteristics: Some pupils have better outcomes than others due to greater family resources, initial endowments of skills, parental inputs and higher quality previous primary school inputs. These attributes may also be correlated with being accepted to a first choice school, via having the resources to move closer to better schools ("selection-by-mortgage").

Preferences: Families have different preferences over schools, and interact differently with the school choice process. The net direction of bias from this source is unclear. While some variation in preferences may be completely idiosyncratic, or related to unobserved variables, much is likely to be explained by observed variables, such as region, religion, ethnicity and socio-economic status.

Area characteristics: the number, quality and variation of the local choice set of schools can influence the probably of getting into a preferred school, as well as subsequent child outcomes via a correlation with regional fixed effects.

2.5.1 OLS and distance matching

In our data, we have detailed information on pupil and family background characteristics, examination outcomes from primary school, primary school attributes, and information about the local area and the choices of schools. Our identification strategy exploits this information, using a conditional independence assumption, motivated by the institutional setting, economic theory, and the prior literature. Our main estimation strategy uses a combination of kernel matching, based on geographical distance, overlaid with regression adjustment, to estimate the average effect of attending a preferred school.

First, we match each pupil who gained a place at their preferred school to a set of control pupils who attend schools in close proximity. We do this using multivariate distance kernel matching on the Euclidean (straight-line) distance.[1] The purpose of this matching to ensure the sample is balanced on geography, such that less reliance is put on capturing geography correctly in the functional form of the regression. Residential location is a choice which may act as a sufficient statistic for unobserved variables—unmeasured wealth, or preferences, for example.

Generally speaking, kernel matching involves the following steps. First, defining a measure of similarity between treated and control units (in this case, straight-line distance). Second, selecting a set of "similar" control units for each treated unit. The control units are selected within a certain bandwidth (radius), with the contribution of the control observations weighted by the shape of a selected kernel function, giving more weight to "closer" observations. Finally, the differences between treated and control outcomes are averaged (over the distribution of treated units' covariates for an average treatment effect on the treated).

In this case, the control units are weighted in inverse proportion to their straight-line distance from the treated unit's school, using an Epachnikov kernel. The bandwidth is selected "optimally" as 1.5 times the 90% quantile of the (non-zero) distances in pair matching with replacement, the default option in Jann (2017) (see Huber *et al.* (2015)).

The matching is performed using multivariate distance matching, specifically matching

[1]Implemented using the software -kmatch- (Jann (2017)), available from "https://ideas.repec.org/c/boc/bocode/s458346.html"

on the Euclidean distance $D_{i,j}$ between two points (i, j), defined by co-ordinates X_{i1}, X_{i2} and X_{j1}, X_{j2} (e.g., Eastings and Northings), computed as follows in Equation 2.1:

$$D_{i,j} = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2} \tag{2.1}$$

Using this geographically matched sample, we use regression-adjustment to estimate the average treatment effect (on the treated) of attending a preferred school. The variables we include in the regression can be loosely grouped into three (overlapping) categories: admissions-relevant variables; variables proxying for preferences over schools; and socio-economic variables. To capture factors which shape admissions probabilities, we adjust for: the presence of resident older siblings; region indicators (government office region); small area fixed effects (Local Authority). We have data on whether the child has been in state-care; however, the proportion is so small that we choose not to match on this.

We also include indicators for whether the child has a certificate of Special Educational Needs—a small group of high needs pupils who may qualify for admission under the "medical and social needs" category. During the time period when our cohort were applying for school, it wasn't required by law to admit pupils with a SEN statement as it is now. We also control for have Special education Needs, but not statement, which is a broader group.

Socio-economic variables are included to indirectly capture aspects of both admission probability and preferences, as well as being associated with child outcomes. For example, family income shapes a family's ability to move closer to their preferred school. In this category we control for (i) the child's prior academic test score—their average points score from their KS2 tests completed at the end of primary school, (ii) Urban-Rural indicator, and (iii) average KS2 points score from child's primary school, (iv) family income, parental education and occupation. The number and quality of schools available to families varies by area, which may correlate with the probability of getting into a preferred school and subsequent outcomes.

We construct a variable characterizing the potential schools a child could have attended; a distance-weighted share of Good or Excellent rated schools by OFSTED (the

Office for Standards in Education is the regulatory monitor of school quality based on school inspections), within a 20 kilometre radius of the school (for urban areas) or 50 km for rural areas (as defined by the Office for National Statistics).

We adjust for variables which proxy for the families' preferences (the type of schools they prefer). In this category we have: (i) religious denomination, (ii) ethnicity of the parents, and (iii) the information used to choose a school. Religion and ethnicity capture parents' preference for attending religious schools, and school with children of the same ethnicity. The LSYPE asks about school preferences. One question asks what information the parents used in their school application decision, listing options such as "looked at league tables on the internet", used "Local Authority brochure", and so on.

Using the geographically matched sample and associated weights, we estimate the parameters of a parametric linear model using OLS as outlined in Equation 2.2. Y_{ijk} is the outcome under consideration, for pupil i, in school j in Local Authority k. c_o is a constant. D_{ijk} is the treatment indicator: equal to 1 for pupils who missed out on their first-choice school, and 0 for pupils who attend their preferred school. X_{ijk} is a vector of covariates. LA_k denotes Local Authority fixed effects.

$$Y_{ijk} = c_o + \beta D_{ijk} + X'_{ijk}\gamma + \text{LA}_k + \epsilon_{ijk} \tag{2.2}$$

We consider two initial sets of covariates:

Specification 1: Share of Good/Excellent schools in local area; female; English second language; religion (dummies, five categories); ethnicity; parental education; any siblings?; government office region; urban-rural indicator; special education needs; parental NS-SEC; parental income; information used in choosing a schools (8 dummy variables, e.g. looked at league tables).

Specification 2: Specification 1 plus: Key Stage 2 average points score; average Key Stage score of the primary school they attended; Local Authority fixed effects.

2.5.2 Channels

Attending a preferred school is a bundle of treatments: the preferred school many have smaller class sizes, better teachers, better peers, and so on. We explore a number of potential mechanisms to unpack the treatment effect of attending a preferred school.

The treatment of missing out varies both within and between schools. At any given school, there are some pupils for whom this is their preferred school, and others for whom it is not. First, we note that the coefficient β is informed by comparison between pupils who get in or miss out both between and within schools.

Missing out on first choice school could impact outcomes through two general channels. First, the indirect effect of being exposed to different school attributes at a non-preferred school compared with a preferred school. Second, a direct "match effect": independent of the particular school characteristics, some pupils may do worse than average at a given school, solely due to the fact that it is not their preferred school. Parents choose schools based on headline measures of "school quality", but also on factors specific to their child: parents are likely to have information about their child's suitability for a given school, which is unobserved by the researcher.

To disentangle these two channels, Equation 2.3 adds current school fixed effects, α_{jk}, such that β is now informed only by within-school variation. This is labelled as *Specification 3* (Specification 2 plus current school fixed effects). Looking at the reduction in the size of the coefficient β, in Equations 2.2 and 2.3, indicates what share of the total effect is due to school attributes, as opposed the remaining match effect.

$$Y_{ijk} = c_o + \alpha_{jk} + \beta D_{ijk} + X'_{ijk}\gamma + \mathrm{LA}_k + \epsilon_{ijk} \tag{2.3}$$

In terms of the role of school attributes, we aim to unbundle the nature of the school effect by examining the effect of missing out on a range of school characteristics, as well as engagement in risky behaviours by the child.

2.6 Findings

Summary statistics

Table 2.1 reports the basic characteristics of the sample. Table 2.2 shows mean differences in key outcomes by whether a pupil got in (control group) or missed out (treated group) on their preferred school. Some of the key characteristics of the pupils associated with gaining a place preferred school include having older siblings, being of White ethnicity, having highly educated parents, begin an owner occupier and being of a Christian faith. Pupils who get in have on average high Key stage 2 scores, while there is little difference in the quality of local schools.

In terms of the attributes of schools themselves, Table 2.4 shows the attributes of preferred and non-preferred schools. Preferred schools tend to have a higher share of White pupils, have a smaller share of FSM pupils, higher levels of achievement in GCSE, higher OFTSED inspection outcomes, and higher value-added.

Table 2.3 shows the raw differences in pupil outcomes by whether they attend their preferred school. Pupils who miss out tend to be more likely to engage in risky behaviours, such as truancy and trying cannabis. Pupils who miss out have, on average, significantly poorer academic outcomes—for instance a lower Key Stage 3 average points score and a lower probability of having 5 or more GCSE passes. In terms of longer run outcomes, there are significant differences observed in a number of outcomes measured at age 25 years. Pupils who do not attend their preferred school are 7 percentage points more likely to drop out of university compared with their counterparts who get in, 4 percentage points less likely to attend a Russell Group university, and a have a lower annual income by -0.08 log-points.

2.6.1 OLS and distance matching

Turning to the regression findings, Table 2.5 presents the effects of missing out on a preferred school on short-run academic outcomes. Specification 1 is the most basic model, including only demographic and socio-economic covariates. Specification 2 adds controls for prior attainment, and Local Authority fixed effects. Specification 3 adds current school fixed effects, to assess the extent to which the treatment effects are comprised of within or between school variation. Missing out on a preferred school could impact outcomes via an indirect effect of attending a poorer quality school. A second channel, comparing pupils who get in or missing out within a given school, is a direct effect of "attending a school you do not want to be attending".

Specifications 1 and 2 in Table 2.5 show little evidence for negative impacts of missing out on short-run academic outcomes. The effect sizes are small and statistically insignificant. The exception is staying on at school at age 16 years to study for A-levels, which is reduced by 2.4 percentage points (or 7% from the control group mean) and is statistically significant at the 10% level. However, this effect disappears after adjusting for school fixed effects, suggesting the effect is likely explained by between- rather than within-school variation. In Specification 3, which includes school fixed effects, there is now a positive coefficient on Key Stage 3 points which is statistically significant at the 10% level. However, it is a very small effect size (0.28 points).

Table 2.6 reports on longer run academic outcomes—attending university at age 20 years, attending an "elite" Russell Group university at age 20 years, gaining a first or 2:1 degree class by age 25 years, and dropping out of university by age 25 years. Considering Specifications 1 and 2, again there is little evidence for detrimental effects on these outcomes. The exception is five-year drop out rates, where dropping out is increased by 5 percentage points among those who miss out (30% increase on the control group mean). This is a large effect size, and it is statistically significant at the 10% level in Specification 1—although not statistically significant in Specification 2. Specification 3 adds school fixed effects—here the effect on dropping out is reduced to 2 percentage points, and is again statistically insignificant.

Finally, Table 2.7 considers a broader set of outcomes: mental ill-health, fertility, contact with the criminal justice system and (log) income. In Specifications 1 and 2, pupils who miss out on attending a preferred school have poorer mental health at age 25 years (about 10% of a standard deviation), statistically significant at the 5% level. Pupils who miss out have reduced income by approximately 1.4% (on a base of about 25,000 GBP per year), which is statistically significant at the 5% level. There is little evidence for any effects on fertility or crime.[2]

Specification 3 adds school fixed effects. In this specification, the negative effect on income remains significant at the 5% level. The effect on mental health is now larger, and statistically significant at the 5% level. This suggests that the between-school variation in missing out on a preferred school has a smaller negative impact on mental health compared with the within-school effect. One interpretation is that parents know something idiosyncratic about where the child will do best—which isn't explained by league tables and other attributes which vary between schools.

2.6.2 Channels

We explore the role of a number of intermediate variables, including alcohol consumption, drugs and mental well-being, which could be plausible variables driving the results. Table 2.8 shows the effects of missing out on these intermediate outcomes. While there is little evidence of an effect on drinking alcohol, playing truant or the GHQ, missing out significantly increases the probability of trying cannabis and increases the risky behaviours score. For example, in Specifications 1 and 2, missing out increases the probability of trying cannabis by 2.4 and 2.3 percentage points, respectively. These effects are statistically significant at the 5% level. Missing out increases the mean risky behaviours score by 0.126 and 0.127 points in Specifications 1 and 2, respectively. These effects are statistically significant at the 5% level. After adjusting for school fixed effects, in Specification 3, the effects on the risky behaviours score become insignificant. The effects on trying cannabis remain, with a similar effect size of 2.4 percentage points.

2.6.3 Heterogeneity

Previous literature has documented gender differences in the effects of school lotteries (Deming *et al.* , 2014). There is also a burgeoning literature assessing gender differences

[2] These results are robust to corrections for multiple hypothesis testing.

in the elasticity to childhood disadvantage (see Autor *et al.* (2016); Brenøe & Lundberg (2018)). These findings motivate an assessment of whether the effects of missing out on a preferred differ by gender. Tables 2.15 and 2.16 show the findings among boys and girls separately. The detrimental effects of mental health are more pronounced among girls, and the income effect more pronounced among boys, although the differences between boys and girls is not statistically significant.

In Tables 2.9, 2.10 and 2.11, we look separately by whether deferred acceptance or first-preference was the dominant assignment mechanism in the child's Local Authority in 2001. We find that the detrimental effects on mental health are more pronounced in areas which used the first-preference system. Table 2.14 shows that these areas are not different based on KS2 attainment or school quality, suggesting that these differences are not explained by FPF areas being less affluent or some other explanation. There is no consistent difference by area in the academic outcomes considered.

While there are small differences in school quality between those who get in and miss out, between DA and FPF areas, the biggest factor appears to commuting distance. The difference in the distance travelled between those who get in and miss out is much larger in FPF areas (those who miss out in an FPF area travel 3.2km further than those who get in, compared with a difference of 0.4km in DA areas). This suggests that pupils who miss out are being allocated a school possibly much further than they would prefer to travel, and may also be a poorer match on other unobservable dimensions. This may explain why the negative effects on mental health are worse in FPF areas.

2.7 Conclusions

This paper examines the long run effects of missing out on a place at a preferred secondary school in England. Employing rich cohort data confidentially linked to administrative data on education outcomes, we compare a range of outcomes between pupils who get into their preferred school and those who miss out. Our empirical strategy leverages features of the institutional setting, literature on school choice and rich administrative data to make a credible case for a selection-on-observed variable identification assumption.

We contribute to a growing literature assessing the long run effects of attending schools which are preferred by parents, providing the first evidence on this topic for state schools in England. During our time period of interest, two mechanisms to allocate school places given preferences and capacities were in operation—a Deferred-Acceptance and a 'first-preference' mechanism—which have been studied extensively in the mechanism design literature. This allows us to estimate separate effect for pupils exposed to each type of assignment mechanism.

In line with much of the international literature, the results do not reveal strong evidence for detrimental effects of missing out on a preferred school on short run academic outcomes. However, those who miss out are more likely to drop out of high school, more likely to drop out of university, have reduced wages and poorer mental health at age 25 years. Increased engagement in risky behaviours, such as drug use and truancy, represent a plausible channel. The negative findings in our paper are more pronounced in areas which used a manipulable mechanism to assign school places (which has since been outlawed).

An important note about the interpretation of these findings is that we are comparing the outcomes of children who look similar observationally similar, and have a similar choice set of schools. From this very specific comparison, it would be incorrect to conclude that "schools don't matter" for academic outcomes.

While the majority of families live within a reasonable commuting distance of a "good" school—at least in terms of headline academic attainment—many of these families are unlikely to gain a place due to rationing by proximity. In this analysis, we take residential location choice and the choice set of schools as fixed—while in reality these parameters can be changed by policy. For instance, changing admission probabilities for families by altering the oversubscription criteria, so it is not so reliant on distance, could conceivably have a large impacts on the school parents apply for and subsequently child outcomes.

A second broader issue, which we intend to investigate in further work, is the extent to which families reveal their true preferences, and the potential welfare loss from not apply to a "truly preferred" school. Strategic families trade-off the probability of acceptance with their true preferences, and list an"acceptable" school as their first choice, rather than their truly preferred school.

A related point is that if some proportion of families are playing it safe with their choices, and we still observe negative consequences of missing out, our estimates could be conceivably seen as a lower bound on an estimate of missing out in a world where if all parents revealed their true preferences (i.e., more ambitious choices of popular schools).

Further work will examine a range of complementary estimation methods (combining geographic and propensity score matching), alternative identification assumptions (employing changes in the local school-age population as instrumental variables) and extended subgroup analyses (by region, prior ability and ethnic group).

Overall, our findings have a number of suggestive policy implications. First, which school a child attends shapes their broader outcomes such as mental health and engagement in risky behaviours, which perhaps should be considered alongside traditional measures of school quality. Second, the nature of the assignment mechanism and how parents understand and engage with it is important for ensuring equal access to schools. Finally, getting into a preferred school matters for important child outcomes. In this light, the oversubscription criteria which act as the gatekeeper to popular schools—determining who gets in and who misses out—is an area ripe for innovation to ensure fair access to schools.

2.8 Tables

Table 2.1: Summary statistics for key variables

	Mean	SD	N
In LA care	0.01	0.08	15,344
Support for SEN	0.13	0.34	15,437
Statement of SEN	0.04	0.21	15,488
Any resident older siblings?	0.86	0.35	15,652
Female	0.49	0.50	15,431
Main parent's higher qualification			
Higher ed.	0.25	0.43	15,087
A-levels, A-C GCSE	0.14	0.35	15,087
Lower GCSE	0.39	0.49	15,087
Other qualification	0.02	0.13	15,087
No qualification	0.20	0.40	15,087
Benefit receipt	0.39	0.49	15,508
Housing tenure			
Owner occupier	0.71	0.45	15,582
Renting from council/LA	0.22	0.41	15,582
Private renter/Other	0.07	0.25	15,582
Religion of main parent			
None	0.24	0.42	15,485
Christian	0.67	0.47	15,485
Buddhist, Hindu, Jewish	0.04	0.20	15,485
Muslim	0.06	0.23	15,485
Ethnicity of main parent			
White	0.88	0.33	15,604
Indian	0.03	0.16	15,604
Pakistani	0.02	0.15	15,604
Bangladeshi	0.01	0.10	15,604
Black African, Black Caribbean	0.03	0.18	15,604
Other, Mixed	0.03	0.18	15,604

Notes: Weighted using the Wave 8 survey weights.

Table 2.2: Summary statistics for pupils (covariates)

	Missed out	Got in	Difference	p-value
Female	0.51	0.49	0.01	0.25
Statement of SEN	0.06	0.04	0.02	0.02
Support for SEN	0.12	0.12	0.00	0.70
Any resident older siblings?	0.79	0.84	-0.05	0.00
In LA care	0.01	0.00	0.01	0.03
Ethnicity of main parent				
White	0.78	0.89	-0.11	0.00
Indian	0.04	0.02	0.02	0.00
Pakistani	0.03	0.02	0.01	0.02
Bangladeshi	0.01	0.01	0.00	0.64
Black African, Black Caribbean	0.08	0.03	0.06	0.00
Other, Mixed	0.06	0.03	0.03	0.00
Benefit receipt	0.49	0.41	0.08	0.00
Main parent's higher qualification				
Higher ed.	0.21	0.24	-0.03	0.01
A-levels, A-C GCSE	0.13	0.15	-0.02	0.08
Lower GCSE	0.39	0.40	-0.01	0.44
Other qualification	0.03	0.02	0.00	0.44
No qualification	0.25	0.19	0.05	0.00
Housing tenure				
Owner occupier	0.61	0.71	-0.11	0.00
Renting from council/LA	0.31	0.22	0.09	0.00
Private renter/Other	0.08	0.06	0.02	0.04
Religion of main parent				
None	0.25	0.23	0.02	0.22
Christian	0.61	0.68	-0.06	0.00
Buddhist, Hindu, Jewish	0.06	0.04	0.02	0.00
Prior attainment and choice set				
KS2 average points score	26.68	27.09	-0.40	0.00
Primary school average KS2	184.95	197.47	-12.52	0.00
Share of schools in local area rated as Very Good	0.27	0.26	0.01	0.05

Notes: Weighted using the Wave 8 survey weights, p-values are computed from robust standard errors clustered by school (the Primary Sampling Unit)

Table 2.3: Summary statistics for pupils (outcomes)

	Missed out	Got in	Difference	p-value
Had a drink?	0.90	0.93	-0.02	0.06
Played truant?	0.20	0.15	0.05	0.00
Ever tried Cannabis	0.11	0.09	0.02	0.01
GHQ (wave 2)	1.73	1.68	0.05	0.50
Risky behaviours	1.10	0.95	0.15	0.00
KS3	32.96	33.92	-0.96	0.00
GCSE	0.58	0.62	-0.03	0.02
Any UCAS points?	0.30	0.36	-0.06	0.00
UCAS pts	214.25	221.72	-7.47	0.08
GHQ (wave 8)	2.59	2.26	0.33	0.01
Whether has Children	0.23	0.20	0.03	0.06
Crime	0.04	0.03	0.01	0.18
Ln(Annual income)	9.56	9.64	-0.08	0.00
At Uni (Age 20)	0.48	0.48	0.01	0.76
Russell group (w7)	0.16	0.21	-0.04	0.02
Upper Class	0.69	0.73	-0.04	0.27
Dropped out of university	0.24	0.17	0.07	0.00

Notes: Weighted using the Wave 8 survey weights, p-values are computed from robust standard errors clustered by school (the Primary Sampling Unit)

Table 2.4: Summary statistics for schools

	Preferred school	Non-preferred school	Difference	p-value
Value-added (KS2-KS4)	988.64	978.26	10.37	0.00
% Free School Meals	17.75	26.59	-8.84	0.00
% English first language	84.52	76.46	8.06	0.00
Derived ethnic composition variables				
% White	76.36	65.79	10.57	0.00
% Indian	4.00	4.62	-0.62	0.01
% Pakistani	4.47	6.65	-2.18	0.00
% Bangladeshi	2.54	3.05	-0.50	0.03
% Black Caribbean	1.79	3.98	-2.19	0.00
% Black African	2.07	4.68	-2.61	0.00
% Mixed	2.66	3.72	-1.07	0.00
% Other (inc. Black other & Asian Other)	2.94	4.80	-1.86	0.00
OFSTED rating (satisfactory or up?)				
Satisfactory+	0.74	0.26	0.11	0.01

Notes: Weighted using the Wave 8 survey weights, p-values are computed from robust standard errors clustered by school (the Primary Sampling Unit)

Table 2.5: Short-run academic outcomes

	Key Stage 3	5+ GCSE	Stayed on at school	UCAS pts.
Specification 1:				
ATT	-0.095	-0.007	-0.026	-1.580
(s.e.)	(0.175)	(0.015)	(0.016)	(4.875)
n	8,602	9,093	9,093	4,145
Specification 2:				
ATT	-0.045	-0.007	-0.024	0.001
(s.e.)	(0.091)	(0.016)	(0.016)	(4.917)
n	7,463	7,823	7,823	3,638
Specification 3:				
ATT	0.283	0.006	-0.011	3.518
(s.e.)	(0.150)	(0.020)	(0.018)	(6.593)
n	7,463	7,823	7,823	3,638
Bandwidth (km)	17	17	17	12

ATT: Average Treatment Effect on the Treated; se: bootstrapped standard error of ATT (with 100 replications); *Specification 1* covariates: Share of Good/Excellent schools in local area; female; English second language; religion; ethnicity; parental education; any siblings?; government office region; urban-rural indicator; special education needs; parental NS-SEC; parental income; information used in choosing a schools. *Specification 2* covariates: Specification 1 plus: Key Stage 2 average points score; average Key Stage score of the primary school they attended; Local Authority fixed effects. *Specification 3* covariates: Specification 2 plus current school fixed effects.

Table 2.6: Long-run academic outcomes

	Started University	Russell group	First or 2:1	Dropped out of university?
Specification 1:				
ATT	-0.015	0.003	-0.033	0.069
(s.e.)	(0.018)	(0.024)	(0.055)	(0.034)
n	6,149	2,934	1,011	2,000
Specification 2:				
ATT	-0.003	-0.002	0.011	0.050
(s.e.)	(0.021)	(0.028)	(0.064)	(0.037)
n	5,354	2,559	884	1,728
Specification 3:				
ATT	0.026	0.021	0.036	0.020
(s.e.)	(0.025)	(0.028)	(0.097)	(0.043)
n	5,354	2,559	884	1,728
Bandwidth (km)	9	8	18	9

ATT: Average Treatment Effect on the Treated; se: bootstrapped standard error of ATT (with 100 replications); *Specification 1* covariates: Share of Good/Excellent schools in local area; female; English second language; religion; ethnicity; parental education; any siblings?; government office region; urban-rural indicator; special education needs; parental NS-SEC; parental income; information used in choosing a schools. *Specification 2* covariates: Specification 1 plus: Key Stage 2 average points score; average Key Stage score of the primary school they attended; Local Authority fixed effects. *Specification 3* covariates: Specification 2 plus current school fixed effects.

Table 2.7: Long-run wider outcomes

	Mental health	Fertility	Crime	Ln(Annual income)
Specification 1:				
ATT	0.318	0.009	0.003	-0.006
(s.e.)	(0.149)	(0.018)	(0.010)	(0.006)
n	4,654	4,848	4,848	4,848
Specification 2:				
ATT	0.334	0.018	0.006	-0.014
(s.e.)	(0.165)	(0.022)	(0.010)	(0.006)
n	4,048	4,210	4,210	4,210
Specification 3:				
ATT	0.540	-0.001	0.000	-0.012
(s.e.)	(0.207)	(0.023)	(0.014)	(0.008)
n	4,048	4,210	4,210	4,210
Bandwidth (km)	11	11	11	11

ATT: Average Treatment Effect on the Treated; se: bootstrapped standard error of ATT (with 100 replications); *Specification 1* covariates: Share of Good/Excellent schools in local area; female; English second language; religion; ethnicity; parental education; any siblings?; government office region; urban-rural indicator; special education needs; parental NS-SEC; parental income; information used in choosing a schools. *Specification 2* covariates: Specification 1 plus: Key Stage 2 average points score; average Key Stage score of the primary school they attended; Local Authority fixed effects. *Specification 3* covariates: Specification 2 plus current school fixed effects.

Table 2.8: Mechanisms: early risky behaviours

	Had a drink?	Played truant?	Ever tried Cannabis	GHQ	Risky behaviours
ATT	-0.006	0.019	0.024	-0.032	0.126
(s.e.)	(0.016)	(0.013)	(0.010)	(0.076)	(0.059)
n	3,615	8,407	8,744	8,583	8,157
ATT	0.000	0.018	0.023	-0.054	0.127
(s.e.)	(0.020)	(0.014)	(0.011)	(0.078)	(0.050)
n	3,223	7,290	7,558	7,429	7,044
ATT	0.002	0.008	0.024	-0.034	0.085
(s.e.)	(0.020)	(0.017)	(0.012)	(0.114)	(0.061)
n	3,223	7,290	7,558	7,429	7,044
Bandwidth (km)	16	17	17	15	17

ATT: Average Treatment Effect on the Treated; se: bootstrapped standard error of ATT (with 100 replications); *Specification 1* covariates: Share of Good/Excellent schools in local area; female; English second language; religion; ethnicity; parental education; any siblings?; government office region; urban-rural indicator; special education needs; parental NS-SEC; parental income; information used in choosing a schools. *Specification 2* covariates: Specification 1 plus: Key Stage 2 average points score; average Key Stage score of the primary school they attended; Local Authority fixed effects. *Specification 3* covariates: Specification 2 plus current school fixed effects.

Table 2.9: Short-run outcomes (by exposure to assignment mechanism)

	KS3	GCSE	Stayed on	UCAS pts.
Specification 1		*“First-preference-first” areas*		
ATT	-0.225	-0.017	-0.024	-13.482
(s.e.)	(0.479)	(0.027)	(0.033)	(13.158)
n	2,294	2,410	2,410	1,075
Specification 2				
ATT	-0.268	-0.027	-0.059	-17.996
se	(0.274)	(0.032)	(0.033)	(16.208)
n	1,919	2,005	2,005	918
Specification 3				
ATT	0.002	-0.005	-0.020	-15.431
se	(0.259)	(0.032)	(0.035)	(14.993)
n	1,919	2,005	2,005	918
Bandwidth (km)	17	17	17	12
Specification 1		*“Deferred Acceptance” areas*		
ATT	-0.051	-0.003	-0.026	1.053
(s.e.)	(0.202)	(0.018)	(0.013)	(5.443)
n	6,308	6,683	6,683	3,070
Specification 2				
ATT	-0.008	-0.001	-0.015	2.367
(s.e.)	(0.122)	(0.017)	(0.015)	(5.077)
n	5,544	5,818	5,818	2,720
Specification 3				
ATT	0.341	0.009	-0.006	6.703
(s.e.)	(0.131)	(0.016)	(0.021)	(6.786)
n	5,544	5,818	5,818	2,720
Bandwidth (km)	17	17	17	12
p-value (1)	0.734	0.660	0.970	0.312
p-value (2)	0.405	0.497	0.224	0.230
p-value (3)	0.215	0.695	0.714	0.215

ATT: Average Treatment Effect on the Treated; se: bootstrapped standard error of ATT (with 100 replications); *Spec. 1* covariates: Share of Good/Excellent schools in local area; female; English second language; religion; ethnicity; parental education; any siblings?; government office region; urban-rural indicator; special education needs; parental NS-SEC; parental income; information used in choosing a schools. *Spec. 2* covariates: Specification 1 plus: Key Stage 2 average points score; average Key Stage score of the primary school they attended; Local Authority fixed effects. *Spec. 3* covariates: Specification 2 plus current school fixed effects.

Table 2.10: Long-run outcomes (by exposure to assignment mechanism)

	At Uni (Age 20)	Russell group (w7)	First or 2:1	Drop out
Specification 1		*"First-preference-first" areas*		
ATT	-0.015	0.010	-0.153	0.135
(s.e.)	(0.045)	(0.053)	(0.206)	(0.091)
n	1,645	707	279	518
Specification 2				
ATT	-0.024	-0.038	-0.145	0.054
(s.e.)	(0.046)	(0.111)	(1.496)	(0.388)
n	1,390	595	240	436
Specification 3				
ATT	-0.010	-0.004	-0.144	-0.006
(s.e.)	(0.046)	(0.139)	(14.214)	(1.595)
n	1,390	595	240	436
Bandwidth	9	8	20	9
Specification 1		*"Deferred Acceptance" areas*		
ATT	-0.012	0.002	-0.014	0.066
(s.e.)	(0.025)	(0.027)	(0.057)	(0.035)
n	4,504	2,227	732	1,482
Specification 2				
ATT	0.006	-0.004	0.061	0.057
(s.e.)	(0.020)	(0.029)	(0.071)	(0.048)
n	3,964	1,964	644	1,292
Specification 3				
ATT	0.038	0.022	0.076	0.027
(s.e.)	(0.026)	(0.027)	(0.129)	(0.052)
n	3,964	1,964	644	1,292
Bandwidth (km)	9	8	20	9
p-value (1)	0.948	0.895	0.502	0.489
p-value (2)	0.566	0.761	0.890	0.993
p-value (3)	0.354	0.851	0.988	0.983

ATT: Average Treatment Effect on the Treated; se: bootstrapped standard error of ATT (with 100 replications); *Spec. 1* covariates: Share of Good/Excellent schools in local area; female; English second language; religion; ethnicity; parental education; any siblings?; government office region; urban-rural indicator; special education needs; parental NS-SEC; parental income; information used in choosing a schools. *Spec.2* covariates: Specification 1 plus: Key Stage 2 average points score; average Key Stage score of the primary school they attended; Local Authority fixed effects. *Spec. 3* covariates: Specification 2 plus current school fixed effects.

Table 2.11: Long-run wider outcomes (by exposure to assignment mechanism)

	Poor mental health	Fertility	Crime	Ln(Annual income)
Specification 1		*"First-preference-first" areas*		
ATT	0.706	0.037	0.015	-0.006
(s.e.)	(0.350)	(0.043)	(0.023)	(0.016)
n	1,307	1,352	1,352	1,352
Specification 2				
ATT	0.987	0.099	0.036	-0.034
(s.e.)	(0.480)	(0.047)	(0.029)	(0.016)
n	1,100	1,136	1,136	1,136
Specification 3				
ATT	1.294	0.054	0.028	-0.025
(s.e.)	(0.478)	(0.059)	(0.030)	(0.019)
n	1,100	1,136	1,136	1,136
Bandwidth (km)	7	9	9	9
Specification 1		*"Deferred Acceptance" areas*		
ATT	0.232	0.001	0.002	-0.010
(s.e.)	(0.175)	(0.021)	(0.012)	(0.008)
n	3,347	3,496	3,496	3,496
Specification 2				
ATT	0.221	0.006	0.002	-0.013
(s.e.)	(0.214)	(0.024)	(0.014)	(0.008)
n	2,948	3,074	3,074	3,074
Specification 3				
ATT	0.450	-0.011	-0.004	-0.011
(s.e.)	(0.231)	(0.034)	(0.013)	(0.009)
n	2,948	3,074	3,074	3,074
Bandwidth (km)	7	9	9	9
p-value (1)	0.238	0.413	0.626	0.849
p-value (2)	0.102	0.109	0.277	0.310
p-value (3)	0.089	0.319	0.329	0.535

ATT: Average Treatment Effect on the Treated; se: bootstrapped standard error of ATT (with 100 replications); *Spec.1* covariates: Share of Good/Excellent schools in local area; female; English second language; religion; ethnicity; parental education; any siblings?; government office region; urban-rural indicator; special education needs; parental NS-SEC; parental income; information used in choosing a schools. *Spec. 2* covariates: Specification 1 plus: Key Stage 2 average points score; average Key Stage score of the primary school they attended; Local Authority fixed effects. *Spec.* covariates: Specification 2 plus current school fixed effects.

Appendix 2.A Descriptive statistics

Table 2.12: Structure of LSYPE

Wave	Fieldwork start	Fieldwork end	Response rate	N
1	30 March 2004	18 Oct. 2001	74	15,770
2	18 April 2005	18 Sept. 2005	86	13,539
3	21 April 2006	28 Sept. 2006	92	12,439
4	12 June 2007	14 Oct. 2007	92	11,801
5	3 June 2008	28 Oct. 2008	88	10,430
6	12 May 2009	14 Oct. 2009	87	9,799
7	18 May 2010	12 Oct. 2010	90	8,682
8	25 August 2015	25 August 2016	51	7,707

Notes: The eligible sample for the Wave 8 sweep comprises all original sample members: the sample identified at Wave 1, and the subsequent boost sample in wave 4. The eligible sample in the previous waves comprises *only those who responded to the previous wave*, leading to a monotone attrition and the appearance of higher response rates from Wave 1 to 7. Alternative response rate calculations are computed in: Collingwood, A., Cheshire, H., Nicolaas, G., D'Souza, J., Ross, A., Hall, J., Armstrong, C., Prosser, A., Green, R., Collins, D., Gray, M., and McNaughton Nicholls, C. (2010). *A review of the Longitudinal Study of Young People in England (LSYPE): recommendations for a second cohort.* DfE Research Report DFE-RR048, Department for Education/NatCen.

Appendix 2.B Timeline of pupil events

Table 2.13: Timeline of pupil events: ages and stages

School year		LSYPE	Year	Age
	Key stage 2 (Primary school)			
Sept. '00 - Sept. '01			Year 6	10/11
	Key stage 3 (Secondary school)			
Sept. '01 - Sept. '02	Start school		Year 7	11/12
Sept. '02 - Sept. '03			Year 8	12/13
Sept. '03 - Sept. '04		Wave 1	Year 9	13/14
	Key stage 4 (Secondary school)			
Sept. '04 - Sept. '05		Wave 2	Year 10	14/15
Sept. '05 - Sept. '06		Wave 3	Year 11	15/16
	Key stage 5 (Secondary school, college, further education)			
Sept. '06 - Sept. '07		Wave 4	Year 12	16/17
Sept. '07 - Sept. '08		Wave 5	Year 13	17/18
Sept. '08 - Sept. '09		Wave 6		18/19
Sept. '09 - Sept. '10		Wave 7		19/20

Appendix 2.C Heterogeneous treatment effects

Table 2.14: Means of key variables by missing out and area type

	Deferred acceptance		
	Got in	Missed out	Difference
KS2 points	27.2	26.9	-0.3
Primary school average KS2	191.1	178.7	-12.4
Share of good local schools	0.3	0.3	0.0
Distance to current school (km)	3.1	3.5	0.4
Value-added	989	979	10
	First preference		
KS2 points	27.3	27.0	-0.3
Primary school average KS2	198.9	187.9	-11.0
Share of good local schools	0.3	0.3	0.0
Distance to current school (km)	3.9	7.1	3.2
Value-added	989	981	8

Notes: This table reports on means of various characteristics by first-preference vs deferred-acceptance areas. The means are weighted using the Wave 8 survey weights.

The tables in this section summarise the effects of missing out on a first choice school on a range of child outcomes. The p-value section (p-value (1), (2) and (3) shows the p-values from tests of whether the effect of missing out on a first choice school differs by gender, for each specification and outcome (the tests all fail to detect gender differences).

Table 2.15: Short-run outcomes (by gender)

	KS3	GCSE	Stayed on	A-levels	UCAS pts
			Girls		
Specification 1					
ATT	-0.150	0.022	-0.017	0.016	0.655
(s.e.)	(0.239)	(0.016)	(0.020)	(0.033)	(5.839)
n	4,290	4,515	4,515	2,238	2,232
Specification 2					
ATT	-0.105	0.004	-0.024	-0.002	2.795
(s.e.)	(0.131)	(0.022)	(0.026)	(0.027)	(6.469)
n	3,746	3,910	3,910	1,972	1,967
Specification 3					
ATT	0.053	0.013	-0.005	0.049	14.086
(s.e.)	(0.161)	(0.023)	(0.027)	(0.053)	(10.033)
n	3,746	3,910	3,910	1,972	1,967
Bandwidth (km)	21	17	17	12	12
			Boys		
Specification 1					
ATT	0.055	-0.024	-0.029	-0.019	-6.553
(s.e.)	(0.246)	(0.020)	(0.023)	(0.037)	(8.964)
n	4,312	4,578	4,578	1,919	1,913
Specification 2					
ATT	-0.012	-0.017	-0.027	-0.012	-11.813
(s.e.)	(0.200)	(0.021)	(0.027)	(0.052)	(9.533)
n	3,717	3,913	3,913	1,677	1,671
Specification 3					
ATT	0.418	-0.005	-0.030	-0.015	-13.382
(s.e.)	(0.216)	(0.025)	(0.027)	(0.052)	(10.420)
n	3,717	3,913	3,913	1,677	1,671
Bandwidth (km)	21	17	17	12	12
p-value (1)	0.591	0.071	0.651	0.496	0.483
p-value (2)	0.680	0.475	0.922	0.866	0.202
p-value (3)	0.155	0.598	0.484	0.419	0.043

Notes: ATT: Average Treatment Effect on the Treated; se: bootstrapped standard error of ATT (with 100 replications); *Specification 1* covariates: Share of Good/Excellent schools in local area; female; English second language; religion; ethnicity; parental education; any siblings?; government office region; urban-rural indicator; special education needs; parental NS-SEC; parental income; information used in choosing a schools. *Specification 2* covariates: Specification 1 plus: Key Stage 2 average points score; average Key Stage score of the primary school they attended; Local Authority fixed effects. *Specification 3* covariates: Spec. 2 plus current school fixed effects.

Table 2.16: Long run outcomes (by gender)

	GHQ	Fertility	Crime	Ln(Annual income)
		Girls		
Specification 1				
ATT	0.388	-0.006	0.005	-0.001
(s.e.)	(0.187)	(0.026)	(0.008)	(0.009)
n	2,596	2,701	2,701	2,701
Specification 2				
ATT	0.458	0.011	0.007	-0.008
(s.e.)	(0.316)	(0.032)	(0.009)	(0.011)
n	2,264	2,355	2,355	2,355
Specification 3				
ATT	0.593	0.006	0.008	-0.005
(s.e.)	(0.312)	(0.040)	(0.010)	(0.015)
n	2,264	2,355	2,355	2,355
Bandwidth (km)	28	28	28	28
		Boys		
Specification 1				
ATT	0.237	0.035	0.002	-0.013
(s.e.)	(0.233)	(0.029)	(0.021)	(0.009)
n	2,058	2,147	2,147	2,147
Specification 2				
ATT	0.312	0.035	0.011	-0.027
(s.e.)	(0.280)	(0.031)	(0.025)	(0.011)
n	1,784	1,855	1,855	1,855
Specification 3				
ATT	0.360	-0.033	-0.002	-0.023
(s.e.)	(0.360)	(0.035)	(0.029)	(0.015)
n	1,784	1,855	1,855	1,855
Bandwidth (km)	28	28	28	28
p-value (1)	0.611	0.344	0.902	0.299
p-value (2)	0.757	0.612	0.892	0.246
p-value (3)	0.638	0.470	0.725	0.417

Notes: ATT: Average Treatment Effect on the Treated; se: bootstrapped standard error of ATT (with 100 replications); *Specification 1* covariates: Share of Good/Excellent schools in local area; female; English second language; religion; ethnicity; parental education; any siblings?; government office region; urban-rural indicator; special education needs; parental NS-SEC; parental income; information used in choosing a schools. *Specification 2* covariates: Specification 1 plus: Key Stage 2 average points score; average Key Stage score of the primary school they attended; Local Authority fixed effects. *Specification 3* covariates: Specification 2 plus current school fixed effects.

Table 2.17: Risky behaviours outcomes (by gender)

	Had a drink?	Played truant?	Ever tried Cannabis	GHQ	Risky behaviours
			Girls		
Specification 1					
ATT	-0.004	0.027	0.024	-0.085	0.141
(s.e.)	(0.026)	(0.016)	(0.012)	(0.120)	(0.073)
n	1,818	4,186	4,352	4,265	4,006
Specification 2					
ATT	0.002	0.034	0.032	-0.082	0.205
(s.e.)	(0.025)	(0.019)	(0.017)	(0.163)	(0.092)
n	1,619	3,650	3,784	3,712	3,479
Specification 3					
ATT	-0.012	0.034	0.036	-0.038	0.209
(s.e.)	(0.035)	(0.025)	(0.019)	(0.177)	(0.073)
n	1,619	3,650	3,784	3,712	3,479
Bandwidth (km)	16	17	17	15	24
			Boys		
Specification 1					
ATT	-0.017	0.014	0.025	0.025	0.112
(s.e.)	(0.028)	(0.019)	(0.014)	(0.101)	(0.067)
n	1,797	4,221	4,392	4,318	4,151
Specification 2					
ATT	-0.024	0.000	0.022	-0.006	0.049
(s.e.)	(0.025)	(0.020)	(0.017)	(0.112)	(0.072)
n	1,604	3,640	3,774	3,717	3,565
Specification 3					
ATT	0.002	-0.020	0.022	0.019	-0.059
(s.e.)	(0.038)	(0.025)	(0.018)	(0.159)	(0.092)
n	1,604	3,640	3,774	3,717	3,565
Bandwidth (km)	16	17	17	15	24
p-value (1)	0.735	0.582	0.957	0.458	0.785
p-value (2)	0.414	0.225	0.614	0.662	0.182
p-value (3)	0.790	0.147	0.578	0.837	0.027

Notes: ATT: Average Treatment Effect on the Treated; se: bootstrapped standard error of ATT (with 100 replications); *Specification 1* covariates: Share of Good/Excellent schools in local area; female; English second language; religion; ethnicity; parental education; any siblings?; government office region; urban-rural indicator; special education needs; parental NS-SEC; parental income; information used in choosing a schools. *Specification 2* covariates: Specification 1 plus: Key Stage 2 average points score; average Key Stage score of the primary school they attended; Local Authority fixed effects. *Specification 3* covariates: Specification 2 plus current school fixed effects.

Appendix 2.D Longitudinal Study of Young People in England

We employ the Longitudinal Study of Young People in England (LSYPE), a nationally representative birth cohort study which tracks the lives of a cohort of around 15,000 young people in England who were born in 1989/90. The study begins when the children are in Year 9, the second year of secondary school, and follows them until they are aged 25 years (annually until age 19/20, then a gap until age 25/26 years). The LSYPE contains detailed information on school choice—including whether the child attends their preferred school—family background, experiences in school and crucially labour market and university outcomes. This data is confidentially linked to the National Pupil Database (NPD) (linkage rate = 97%), a database of administrative school records, containing test scores and school attributes. From this data we can track the achievement of the LSYPE pupils in high-stakes examinations, e.g., primary school examination results, GCSE and A-levels. The NPD also provides information on the attributes of schools.

The sampling frame comprised all pupils attending maintained schools, independent schools and pupil referral units in England on February 2004. The first 7 waves of the study were funded by the Department for Education (DfE), commissioned to explore the factors shaping educational attainment and transitions out of compulsory schooling. The final wave, at age 25 years, was funded by the Economics and Social Research Council, and management transferred to the Center for Longitudinal Studies, now with a wider remit to explore wider aspects of the transition to the labour market.

The LSYPE has a a complex sampling design, involving clustering and stratification. Clustering involves categorising a population by a grouping variable (e.g., schools), then a number of clusters are selected via some rule, e.g. a random sample, and then cases within those clusters are sampled. One reasons for clustering is to enhance the cost-effectiveness and convenience of data collection. For example, the LSYPE clustered by school. Because the pupil's addresses were not available from Pupil Level Annual School Census (PLASC) at that time, the addresses were collected from schools. Clustering tends to increase the sampling variance; without accounting for the clustered nature of the data tends to over-estimate the precision of an estimate, the standard errors would be too small.

Stratification involves categorization of a population by a grouping variable (e.g., deprived/non-deprived), and then cases are sampled independently within each strata. A common motivation is to oversample in particular strata, in order maintain a sufficient sample size in specific subgroups: LSYPE oversampled ethnic minority pupils and the most deprived schools, for example. More generally, stratification can increase the precision of estimates in subsequent analysis. However, without adjusting for this type of stratification in substantive analyses, such as by down-weighting the oversampled groups, estimates will not necessarily represent the target population.

LSYPE used a two-stage sampling design (first sampling schools, then pupils within those schools). Maintained schools were selected from the Pupil Level Annual Schools Census (PLASC), stratified by deprivation status: deprived schools—those in the top quintile of the distribution of free school meal receipt—were oversampled by a factor of 1.5. Within each deprivation stratum (top quintile vs bottom four quintiles), the schools were then ordered by region (London/not London), then by school admissions policy, before the selection of pupils (this is a way to implicitly stratify on these variables). Independent schools and PRUs were sampled from the School Level Annual Schools Census (SLASC). Independent schools were stratified by percentage of pupils achieving 5 or more A*-C GCSE grades in 2003 within boarding status (i.e. whether or not had any boarding pupils), by gender of pupils (i.e. boys, girls and mixed). PRUs formed a stratum of their own.

The second stage sampled pupils within the selected schools. Home-schooled pupils, pupils in very small schools (fewer than ten pupils for maintained sector, and fewer than 6 for independent schools), and children residing in the UK solely for education purposes, were excluded. Within the maintained sector, pupils were sampled from PLASC. Pupil selection probabilities were dependent on ethnic group (White; Indian; Pakistani; Bangladeshi; Black African; Black Caribbean; and Mixed), and on school selection probabilities. According the User Guide, "The school sampling stage took into account the number of pupils from each [of these] minority groups. Taken together, the school selection probabilities and the pupil selection probabilities ensured that within a deprivation stratum, all pupils within an ethnic group had an equal chance of selection." The pupil selection aimed to sample at least 1,000 pupils in each ethnic group. The average number of pupils sampled per school was 33.25, although the number sampled per school varied according to the ethnic group composition of the school population.

Pupils in independent schools and PRUs were sampled directly from school rolls by interviewers using a sampling program installed on their laptop computers. 33 or 34 Year 9 pupils (33.25 on average) were randomly selected at each independent school or PRU containing 34 or more year 9 pupils. All the Year 9 pupils were selected in the independent schools / PRUs that contained fewer than 34 but more than 5 Year 9 pupils. Of the 892 schools selected in total, 647 schools (73%) co-operated with the study.

Table 2.12 shows the survey structure of the LSYPE. The fieldwork was mainly carried out between April and October, with the bulk occurring toward the end of the academic year, i.e. April to August. However, a minority of the interviews fell in the subsequent academic year. Pupils applying to university would have passed most stages of the process and be holding offers by interview.

The LSYPE achieved cross-sectional responses rates ranging from 51% (in wave 8), to 92% (in Waves 3 and 4). For wave 1 to 7, the sample issued (response rate denominator) at each wave comprised respondents from the immediately preceding wave, who agreed to be re-contacted (rather than all participants identified as the sampling frame in wave 1). The exception is the Wave 4 sample frame also included an ethnic minority boost of six hundred Black African and Black Caribbean young people. This sample was selected from schools who did not co-operate in the initial Wave 1 sampling frame. This boost had a response rate of 59%, adding an additional n=352 participants. Therefore, despite reasonable cross-sectional response rates, over time the sample size was severely reduced. Response rate (2) in Table reports the number of respondents in each wave divided by all Wave 1 respondents.

Wave 8, now managed by CLS, widened the issued sample from not only those who were in wave 7 who agreed to be re-contacted (the procedure in previous waves), but all participants: the 15,770 from wave 1 plus the Wave 4 ethnic boost of 352, to yield a sampling frame of 16,122 cases. A small number asked to op-out, leaving 15, 629 to be contacted for the Wave 8 interview. Of the 15,531 sample issued, 423 were ineligible, and productive interviews were achieved with 7,707 cohort members, yielding a cross-sectional response rate of 51%, and the longitudinal Wave 7 to 8 figure was 69% (i.e. 69% of those responding to Wave 7 also responded to Wave 8).

The interviews in Waves 1 to 4 were carried out face-to-face at the young person's home using computer assisted personal Interview (CAPI). From Wave 5 onward, the survey used mixed-methods (online, telephone and face-to-face interviews). Wave 5 comprised 32% online interviews, 44% telephone interviews, and 12% face-to-face interviews at the participants' homes. From Wave 5 onward, only the young person as interviewed, not the parents. Wave 6 comprised 39% online interviews, 48% telephone interview and 13% face-to-face. The figures for Wave 7 are 40% online, 40% telephone and 18% face-to-face (a further 2% face-to-face interviews were issue but not completed). Wave 8 also deployed a sequential mixed mode strategy, with an online option first (64%), then telephone interview (9%) and finally face-to-face (29%).

The fact that many of the participants in wave 8 had not participated in each wave, due to the sequential sampling method, means that out of the 7,707 Wave 8 participants, many do not have a complete history from participation in all waves. Out of the total sample of 16,122 participants, 33.7% (n=5,426) of all respondents partook in all 8 Waves, 16.7% (n=2,694) had interrupted response –largely due to dropping out somewhere between Wave 1 and 7, and then being re-issued at Wave 8; 49.6% (n=8,002) had a monotone pattern of response, that is, they partook in some number of consecutive waves then permanently dropped out. Another issue is that if you add to the denominator the pupils in the schools which didn't respond in the first wave–then response rate are lower, for example 51% rather than 74% for Wave 1.

Survey drop-out both reduces power, due to the reduced sample size, and can lead to biased analyses, depending how the research questions at hand relates to missingness. Weights are provided with each LSYPE wave to account for account for the complex survey design, and non-response. The Wave 1 weights account for the sampling in the following way: design weights (accounting for the survey design) are computed separately by maintained vs non-maintained schools (independent). 28 out of the 54 independent schools issued with questionnaires responded, yielding 530 pupil responses. The independent school weights were computed to weight the responding schools to match the population figures on on variables found to be associated with non-response: school type (mixed/boy/girls) and London/other regions.

Out of the 838 schools issued in the maintained sector, 646 responded. Maintained sector weights were computed for school non-response, pupil non-response and then calibrated to totals. School non-response was explored by share of non-White pupils, share of pupils with 5+ A*-C GCSE grades, and deprivation status of the school and region, weights were created based on deprivation band and region, which were significant predictors of school response. The pupil non-response weights used gender, ethnicity interacted with region, GCSE attainment. The school and pupil response weights were combined and calibrated to create weights which matched populations totals (ethnicity, region, GCSE outcome and gender, from NPD), and then adjusted to match independent vs maintained population totals.

Weights for subsequent waves were calculated by modeling the probability of continuing from Wave n to Wave $n + 1$. These weights were multiplied by the weight from the previous wave, to account also for the survey design and selection into the previous wave. A similar procedure was carried out at Wave 8 by CLS: response at wave 8 was modeled as a function of family background and young person's characteristics at Wave 1 (including: parental qualification and employment, NS-SEC, marital status, ethnicity and gender of the child); and the young person's characteristics at Wave 7 (including: housing tenure, labour market/education status, ever tried cannabis, interview month and interview mode). Missing data on predictor variables—due to item non-response, or having not participated in Wave 7—was handled via multiple imputation. This non-response weight was then multiplied by the Wave 7 final weights, which accounts for non-response up to Wave 7, and the complex survey design.

Solon *et al.* (2015) classify the situations where weighting is appropriate into three groups: (i) estimating population representative quantities; (2) if missingness or survey design is correlated with the association, or causal relationship in question, and (iii) heterogeneity. In this analysis we use the weights throughout—we are interested initially in describing population quantities, such as how many pupils miss out on a preferred school, in its own right, and (2), it seems plausible that the relationship between missing out and later outcomes could be confounded by survey drop out.

Appendix 2.E Institutional setting

2.E.1 Historical setting

Beginning with Conservative administrations between 1979 and 1997, a number of significant reforms have been introduced which have substantially modified the operation of the market for secondary schools.

The 1988 Education Reform Act introduced a number of important changes increasing the agency of parents and schools, including open-enrolment—allowing parents to apply to any school in England, rather then being constrained by the boundaries of their LEA. It also prevented schools from rejecting pupils for any reason aside from over-capacity. The 1988 Act also initiated a new development in school governance, "grant-maintained status". Grant-maintained schools could manage their own admission arrangements (similar to voluntary-aided schools, which were often church schools). This increase in choice from schools over pupils, and increase in choices available to parents, was a significant landscape of school choice and governance (Committee on Education and Skills, 2004).

A number of concerns emerged at this point (West *et al.* , 2004), including a lack of co-ordination of admission policies across school types, local areas and admission authorities; increased complexity of choosing and applying for a school; potential for "cream-skimming" among grant-maintained and voluntary-aided schools (West *et al.* , 2004). In advance of the 1997 election, the Labour party manifesto highlighted education its first priority, committing to two key themes: limiting selection (from schools) and enhancing (parental) choice through diversity of school types. The commitment to reducing the degree of selection by schools was made clear by David Blunkett MP, the shadow Education Secretary at the 1995 Labour Party conference: *"read my lips: no [more] selection, either by examination or interview, under a Labour government"*. The right of parents to express a preference was strongly affirmed: *"all parents should be offered real choice through good quality schools, each with it's own strengths and individual ethos"* (Committee on Education and Skills, 2004).

The key piece of legislation introduced by Labour, once in Government, was the 1998

School Standards and Frameworks Act (SSF Act). The goal of the SSF Act was to make admissions transparent, objective and allow each child a fair chance of satisfactory place. The SSF Act required the Secretary of State to publish a Code of Practice regarding schools admission—a document providing guidance on best practice in admissions. This and subsequent editions of the School Admission Code contain descriptions of the primary legislation from which they stem, extensive guidance, and reference to statutory responsibilities which must be met, but are not themselves a legal document. The Codes *"signpost the relevant legal provisions but they do not aim to provide definitive guidance on the interpretation of the law: that is a matter for the courts."* (cite: School Admissions Code of Practice, Department for Education and Skills, 2003, para A1).

The guidance associated with the SSF Act was the Code of Practice on School Admissions 2001 (DfE, 2001), enacted on 01 April 1999, and in force for admissions for intakes in September 2000 onward. Schools and admission authorities were required to "have regard to" the indications in the Code. The Code encourage increased uptake of common admission processes: *"LEAs should consider, with other admission authorities, having co-ordinated admission arrangements - including standard application forms and common timetables - for all schools"* (DfE (2001), para. 3.9). Each school has an admissions authority which decides which children are admitted to the school. While for maintained (government controlled) schools, this was generally the Local Authority, for foundation[3] (grant-maintained), voluntary-aided (church schools) and City Technical Colleges, it was typically the school governing body (Williams *et al.* , 2001).

The Code provided guidance on oversubscription criteria, mandating that these criteria and exactly how they are used need to be published, and in be accordance with the law: *"Admission authorities have a fairly wide discretion to determine their own oversubscription criteria provided these criteria are objective, clear, fair, compatible with admissions and equal opportunities legislation."* (DfE (2001), para. 5.2). The Code provided some suggestions for criteria to be used: *"Commonly used and acceptable criteria include sibling links, distance from the school, ease of access by public transport, medical or social*

[3]the 1998 Framework converted grant-maintained schools either to voluntary-controlled schools or, most commonly, foundation schools.

grounds, catchment areas and transfer from named feeder primary schools, as well as parents' ranking of preference." (DfE (2001), para. 3.14). *"Admission authorities should make clear the order of priority in which the criteria will be applied, and how any tie-break decisions will be made."*(DfE (2001), para. 5.3)

After the 2001 Code, the next piece of primary legislation introduced was The Education Act 2002, and the associated School Admissions Code 2003. Some of the key requirements introduced the 2002 Act are: (i) a requirement for LEAs to coordinate admission arrangements between admission authorities and across LEA boundaries; (ii) the statutory creation of admissions forums in each local authority, in which all admissions authorities must participate; (iii) explicit advice (but not a statutory requirement) that priority should be given to children in public care (Committee on Education and Skills, 2004). The School Admissions Code 2007 and School Admissions Code 2010 followed.

Admissions in 2001

The admission process varied both within and between local authorities, and in September 2001 was much more heterogeneous and under-developed compared with the present day arrangements. At its most general level, the admissions process three stages (i) parents submit their preference(s) to the admission authority (or authorities), (ii) the admission authority processes tries to assign families to their preferred schools, and (iii) in the case of oversubscription, applications are ranked based on a set of observable criteria to ration scarce places. In practice, the actual procedures followed vary greatly and are difficult to categorise (see West *et al.* (2004); Williams *et al.* (2001) for surveys of the admission protocol used, and specific examples from LEA brochures).

Once the families' preference rankings have been submitted, a procedure is required to allocate school places, given school capacities and priorities. There are a number of approaches for doing this. The two used in 2001 are (1) the *first-preference* algorithm, also called the Boston mechanism, or immediate-acceptance, and (2) the *equal-preference* algorithm, also called the Gale-Shapley mechanism, or deferred-acceptance. West *et al.* (2004) found that 41% of schools used a 'first-preference' approach to offer places for the

September 2001 intake. Reproducing the helpful example in Cantillon (2017), this section briefly describes the main features of these two approaches as relevant for this paper.

Suppose we have three schools, A, B and C, and to simplify, only one priority group is used in admission, older siblings at the school, and then places are rationed using the tie-breaker of straight line distance from the school. Each school has one place to offer, and there are three students competing for places: Anne, Bob and Chloe. Table 2.18 lists the preference ordering over the schools for Anne, Bob and Chloe. School A is known in the community to be the best school (both Anne and Bob rank it first choice), and is typically oversubscribed. School C is seen as the worst school which is described as a bad outcome (everyone's last choice).

Table 2.18: Anne, Bob and Chloe's preferences over schools, and whether they have siblings at the school

Anne	Bob	Chloe
School A (sibling)	School A	School B
School B	School B	School A
School C	School C	School C

After applying the admission criteria—older siblings and distance—the ranking of the pupils at each school are listed in Table 2.19. Anne's sibling is at School A, so she has priority at School A, but not at school B. Bob lives closer to School A and B than Chloe, and so on.

Table 2.19: Ranking of the pupils at each school, given their preferences and priorities (siblings and distance)

School A	School B	School C
Anne (sibling)	Bob	Anne
Bob	Chloe	Bob
Chloe	Anne	Chloe

The first-preference method starts by each school assigning places to pupils *who put that school as their first choice*, based on the ranking determined by the school's admission criteria—until either all there are no places left, or all students who put that school as first choice are assigned. The next round conducts the same procedure for pupils who put that school as their second-choice, and so on. In this example, Anne is offered a place at School A, because she listed School A as her first choice, and is ranked higher than Bob, who also put School A as his first choice (due to Anne's sibling at School A). Chloe put School B as her first choice, and is assigned place at School B (Bob and Anne are out of this round). Bob is assigned the remaining place at School C—his least desired outcome.

Bob is penalised for revealing his true first choice of School A. Had he put School B, he would have been assigned a place there over Chloe by the distance rule. Hence, while the first-preference method tends to have a a higher share of pupil's getting into their first choice; it also incentivises strategic behaviour. Families have an incentive to trade-off their true preferences with their probability of acceptance. Without this consideration, if families put a school as their first choice that have a very low chance of acceptance into, they may not only fail to secure a place at their preferred school but also fail to be placed at any school they deem acceptable.

The problem is that the preference rankings are *themselves are used as a priority*. In contrast, the 'equal-preference' approach does not use the preference ranking as a priority. In this method, parents submit their preference ranking over schools. The families' application is considered at each school that they listed as a preference without reference to its ranking. In the case of listing a school with no excess demand, the family is offered a place at that school. In the case of having listed an oversubscribed school, the oversubscription criteria are applied and places are offered up to capacity based on the resultant ranking. After receiving offers, families can accept their most preferred school. Once places are freed up from rejections, the place is offered to the next student in the ranking. The process continues until all places have been offered (this process may be carried out by a computer).

In this example, the first stage consists of offering the place at School A and School C to Anne, because Anne is the top of the priority list at these two schools, and offering a place at School B to Bob. Anne rejects School C, because it is not his preferred school.

Bob holds onto his only offer, School B, for now. School A now has no more places to offer. In Stage 2, School C, the only school with a spare place, offers it to Bob–the second down on their list. Bob rejects, and accepts School B which he prefers. Stage 3, School C is the only one with a place left, and offer the place to the next free student in the ranking, Chloe, who holds onto it as her only current option, hoping she may still receive a better offer. However, now all the places are offered, and the process ends and the places are confirmed. [4]

In 2001, while admissions authorities had a statutory obligation to have regard to parental wishes, they had discretion as to what role, or priority, parental preference took in relation to other oversubscription criteria in determining admissions: *"Where parents can express more than one preference, the order of priority by which parents rank their preferences may be given priority over any other means of determining how to allocate places at oversubscribed schools, but that is not a statutory requirement. An LEA must have regard to the guidance in this Code in drawing up its admission policy. It may do this by adopting the parents' order of ranking as one of its criteria, if it regards that as a fair and beneficial way of determining such cases and maximising parental preference; or it may adopt an admissions policy which applies some other criterion, compatible with parents' preferences and the guidance and objectives set out in this Code."* (DfE (2001), para. A.30). In the survey of schools by West *et al.* (2004), 41% of schools used a 'first-preference' system.

This practice of using preference order as a priority was prohibited in the School Admissions Code 2007: [schools must not] *"give priority to children according to the order of schools named as preferences by their parents, including 'first preference first' arrangements;"* (School Admissions Code 2007, 2.16 (b)). *"the 'first preference first' criterion made the system unnecessarily complex to parents"* and *"forces many parents to play an 'admissions game' with their children's future."*(cite: School Code 2007, Foreword, p. 7).

In the nationally representative survey conducted in Flatley *et al.* (2001), about 25% of parents reported that took into account the nature of the over-subscription criteria when

[4]This examples describes the school-proposing variant of DA, which in this specific example gives the same result as the pupil-proposing variant. However, in general they are different and the pupil-proposing variant Pareto dominates the school-proposing variant for pupils. In practise, a mix of these two variants is used across LEAs

considering which schools to apply to. Parents who were familiar with the oversubscription criteria in their preferences tended to be more highly educated, owner occupier and of white ethnic origin:

- families where the mother had a degree, or higher qualification, were three times more likely than those without any qualifications, and approximately twice as likely as those with lower qualifications, to report that they knew how popular schools allocated;

- owner occupiers were about twice as likely to say they knew as were parents who were social renters;

- families where the mother was of white ethnic origin were nearly twice as likely to say they knew as those with a mother of non-white ethnic origin.

These different levels of strategy are reflected in testimony from a school admissions officer interviewed in Flatley *et al.* (2001):

"The problem that we have been unable to overcome is that a significant number of parents fail to understand that their local community school may not be available to them as second preference if they take a gamble by stating a first preference for a voluntary-aided school whose religious adherence they do not share, or a popular community school not very close to where they live. We publish a list of the schools that have been oversubscribed over recent years, but this does not succeed in deterring every applicant from naming one of those schools as second preference. This is not really a disadvantage of the system, but our failure to communicate with parents."

2.E.2 Oversubscription criteria used in September 2001

Regarding pupils with a Statement of Special Education Needs, the 2001 Code specifies that an admission authority has a duty to admit pupils with a certificate of special educational needs *who named that particular school on their certificate*: *"Where a school is named in a statement of special educational needs, the admission authority has a duty* *to admit the child to the school."*(DfE (2001), para. 5.8). It emphasized that pupils who require support for Special Education Needs, but have no certificate, were not to be discriminated against (i.e., due to their potential for extra resources being required to cater for their needs).

Interviews with parents were now specifically prohibited, aside from in order to conduct a religious assessment: *"Schools or admission authorities should not interview parents as any part of the application or admission process. Church schools may carry out interviews, but only in order to assess religious or denominational commitment."*(DfE (2001), para. 5.25).

In 2000, it wasn't required to give "looked-after" children first priority if a school is oversubscribed. In the 2003 Code, it was suggested to give children in care priority, and from 2006, it was a statutory requirement. In 2008, almost all schools (99%) had an admissions criterion relating to children in care compared with only 2% in 2001 West *et al.* (2011).

The 1998 SSF Act allowed secondary schools to use "fair banding" in their admissions, a method aimed at ensuring balanced distribution over ability. Specifically, Section 101 of the SSF Act permits admission authorities to employ arrangements which select children by general ability, only if the arrangements are *"designed to secure that in any year children admitted into a normal year of entry are fully representative of the range of ability amongst children applying to the school for that year of entry (which may well be different from the range of ability nationally), and that no level of ability is substantially over- or under-represented"* (DfE (2001), A.65) Often the 11 plus test is widely taken in the LEA, and the results are also used in comprehensive schools to ensure that the intake is balanced over the test performance distribution. The usual oversubscription criteria are then applied if there are more children falling within a particular band than places allocated to that band.

The 1998 SSF Act allowed the admission authority for a school with a "specialism" (*"particular expertise or facility"*—it does not require the school to be in the formal specialist schools programme) to give priority to up to 10% of pupils who can demonstrate an 'aptitude' in the relevant subject (DfE (2001), A.69). The 10% limit is an overall limit, not per subject (DfE (2001), A.70). The subjects which can be used are: physical education or sport; performing arts; modern languages; design and technology; visual art (DfE (2001), A.71). It is still unlawful to test for "ability" or aptitude in any subject aside from the ones prescribed above.

As mentioned in the previous section, some schools used families' preference ranking in their oversubscription criteria ('first-preference-first').

What do we know about what admissions were actually used in 2001?

Two key surveys of admission policies were conducted during this time period, including a survey of all government-funded schools by West *et al.* (2004), and a DfE-commission survey of schools and parents conducted by Flatley *et al.* (2001) and analyses in Williams *et al.* (2001). Both of these surveys have associated follow-up studies comparing changes in admission processes over time, see West *et al.* (2011) and Coldron *et al.* (2008) respectively.

West *et al.* (2004) conducted a survey of all government-maintained secondary schools for pupils entering in September 2001. West *et al.* (2004) were able to collect data for 95% of secondary schools (*N=3,013*), and summarised the admission information from the admission prospectuses. The key table summarising the frequency of different admission criteria is reproduced in Table 2.20. 13% of schools referenced religious criteria; 73% had regard to "medical or social needs" of the child; 39% referenced "special needs"; 41% used 'first-preference'; 1.7% referenced children in public care. Only 2% of schools reported interviewing pupils; in accordance with the DfE (2001) which excludes interviewing, aside from for religious purposes. The review concluded that, overall, the admissions criteria do not appear to be unfair or be designed to advantage a particular type of student at the expense of others. However, they did highlight the practices of specific types of schools—with more control over their admission criteria—which did appear to try to select certain students.

Williams *et al.* (2001) collected information from the composite prospectuses sent to parents in 1999, for children starting in September 2000, with coverage of 100% LEAS which have secondary schools where pupils transfer from primary at age 11 years (*n=141 LEAs*). Similar to West *et al.* (2004) , they highlight the variation in admissions arrangements and oversubscription within and between LEAs. Williams *et al.* (2001) find a range of different ways in which parents could express their preferences, and categorise them into four loose groups—noting that this categorization is necessarily an oversimplification. The most common category (74.4%) was for parents to submit a ranking of preferences, and the LEAs authority tries—either using a first- or equal-preference approach—to allocate families to a preferred school. However, sometimes there were exceptions to this, including where a first-choice could be overridden if another pupil, who put the school as second choice, had an especially long or unsafe journey or some other exceptional circumstance. In some occasions, parents were offered a second round to submit further preferences, if their first set of preferences were unsuccessful. In the second category (12.8%), LEAs offer a place at a school to a family; and then families can submit this school as their first preference, or alternatively the family can submit a ranked list of different preferred schools. The third category (5.0%), parents submit multiple unranked preferences, and the LEA aims to allocate the family to one of these preferred schools. The final category (7.8%) is sequential: families are invited to submit a single preference, if this proves unsuccessful, a second round of submitting preferences takes place.

Williams *et al.* (2001) present a number of concrete examples of the many different admissions arrangements, including the following:

"The LEA uses a common form that every parent has to complete but if parents are interested in voluntary-aided and foundation schools those schools have their own admission forms and parents complete an additional form for that particular school. There is a two-part form and on one side they choose up to five schools that they are interested in. On the second side, which is confidential to the LEA, parents list an order of preference which only comes into play if parents meet the admission criteria for a place at more than one school. If they satisfy the admission criteria of more than one school the LEA looks

at the order of preferences but this is not shown to the schools. The LEA allocates places for its community schools and has the admission lists for the other voluntary-aided and foundation schools. The schools provide the LEA with a list of children that have been offered places. The LEA identifies those that are going to be offered a place at more than one school, and then rejects the lower preference places. The information will normally be sent to a voluntary/foundation school saying that it has been possible to meet a parent's preference. It is only when parents meet the admission criteria for more than one school that the list of preferences is looked at."

Sources of variation in admission

There are a range of sources of variation and uncertainty about whether one will be admitted to a particular school.

Catchment area policies, while becoming less popular at that time, were used in both general admissions (14.9%) and featured somewhere as an oversubscription criteria in (63.1%) of LEAs Williams *et al.* (2001). Catchment area policies create uncertainty because the catchment area can change over time in unpredictable ways. In the survey by Williams *et al.* (2001), one LEA deployed a flexible catchment area system based on the numbers of children applying to particular secondary schools from particular primary schools over a period of time. Additionally, sibling and catchment criteria can interact; for example, siblings in a catchment area, other children living in a catchment area, siblings living outside a catchment area and, finally other children outside a catchment area. Given random fluctuations in the sibling distribution, this creates some random variation in the effects of the catchment area on the probability of admission. The changing interaction of sibling and distance criteria makes it difficult for parents to predict their chances of gaining a place at a particular school, complicated further by the interaction with banding, where applicable, as evidenced in this testimony from a school admissions officer in the survey by Williams *et al.* (2001):

"Most local children are guaranteed a place but there have been years where a whole band has been filled up with siblings and it changes. On average nearly 50% of our pupils

will be siblings - so that's 120. If they all happen to be in the same band, there's a problem! It's never happened, but I suppose it could be that there's 70 pupils in the top band that have all got siblings".

Distance to school also creates uncertainty, because families do now know in advance whether or not they live close enough to a school in any given year. Although some individual schools, in some LEAs, publish information over several years about the 'cut-off point' for distance, typically it is not clear how near 'near enough' will be the next year (Williams *et al.* , 2001). The first preference system is also a source of uncertainty in the probability of acceptance, because it depends, amongst other things, upon how many parents apply to a school in a given year, which fluctuates year-on-year and cannot be known in advance (Williams *et al.* , 2001).

Outside options

Some areas have grammar schools, which wholly select on ability. How these schools are accounted for in the admissions process varied. The first is where parents submit completely separate preference lists to selective and non-selective schools. When the child's 11 plus results are known, the appropriate list is consulted. The second approach is that parents express their preferences for both selective and non-selective schools on the same form. When the 11 plus results are known, the highest preference non-selective school becomes their effective first preference if they do not pass the 11 plus. The third approach is that parents express their preferences for both selective and non-selective schools on the same form. When the 11 plus results are known, the list is interpreted literally: if a selective school is first preference and a non-selective is second preference, the non-selective school is still counted as a "second preference". This is a disadvantage when parents are unsure a priori of their children's eligibility for the selective school (Flatley *et al.* , 2001).

Table 2.20: Non-exhaustive frequency of secondary schools admissions criteria practice (excluding grammar schools) from West *et al.* (2004)

	%
Siblings	96
Distance	86
Medical/social need	73
Catchment area	61
'First preference'	41
Special educational needs	39
Feeder school	28
Religion	13
Children of employees	9
Difficult journey	6
Children of former pupils	5
Banding	3
'Other faiths'	3
Aptitude in subject area	2
Pupil interview	2
Strong family connection	2
Parent interviews	2

Chapter 3

Primary school preference, child outcomes and parental response

I am grateful for helpful comments from my supervisors Ian Walker and Eugenio Zucchelli. This work was supported by a PhD studentship funded by the ESRC and the Department for Economics at Lancaster University. The initial data preparation for this chapter was undertaken while I was seconded to an ESRC-funded internship at the Department for Work and Pensions Children and Families Unit. Thanks to David Church at the Centre for Longitudinal Studies for providing the distance from home to school variables.

3.1 Introduction

Schooling exerts long-lasting causal effects on a range of outcomes over the lifestyle, including health (Jones *et al.* , 2011), cognition (Banks & Mazzonna, 2012), and labour market outcomes (Harmon & Walker, 1995). Given school performance varies, and places at the best schools are limited, how these places are allocated is potentially important for the distribution of subsequent outcomes.

In England, there are differences in the measured academic performance of different primary schools, and places at the best schools are hotly contested. Many families move house to secure a place at good schools (Hansen, 2014): 18 per cent of parents reported moving house during the two years prior to starting primary school, for reasons to do with their child's education. This preference for good schools is capitalised into house prices. Gibbons & Machin (2003) show that an 1 percentage point increase in the share of local pupils meeting the required academic target increases house prices by 0.67% in England.

While the majority of families are offered a place at their preferred school, some inevitably miss out. Not attending a preferred school may influence outcomes via attending a lower quality school, being exposed to a different peer mix, or a sub-optimal school match for the child. This setting offers an opportunity to credibly assess the effects of a perceived early-life "shock", both on child development and parental responses. For example, parents may respond to this perceived disadvantage by paying for private tuition, which is increasingly prevalent among primary school pupils in England, or helping with homework.

Behavioural responses are especially important in education settings, because the production of educational outcomes involves the interaction of pupils, parents, school teachers and administrator and the policy context. Parental responses to public investment also differ by demographic group: recent evidence showed that by age 11 years, 22% of English primary school students were receiving private tuition, with large variations by ethnicity and social background (Kirby, 2016). These types of interactions could dampen or accelerate policy interventions in unintended ways. Despite the importance of understanding these public-private interactions in educational investments (Albornoz *et al.* , 2018), there are few studies which credibly assess behavioural responses to early-life shocks. Recent papers include parental responses to class size (Fredriksson *et al.* , 2016), responses to month-of-birth (Bernardi & Gratz, 2015) and responses to birth weight and fetal health (Restrepo, 2016; Almond & Mazumder, 2013). These papers report mixed evidence on the presence of parental responses.

Missing out on a preferred school makes a compelling contribution to this literature for a number of reasons. First, the school choice and admission process is widely publicised in England—parents are aware of the importance of getting into the best-suited school for their child, and are likely to see missing out on their preferred school as a disadvantage to their child—perhaps more so than other inputs such as month-of-birth—making it a useful case study. Second, the deterministic nature of the admission rules for primary schools offers a study design especially amenable for credible assessment of the effects of missing out on parental behaviours. Finally, primary school is an especially relevant period to investigate parental responses in particular: in addition to being critical period for skill development, greater investment at primary school may increase a child's chances of entry into a selective secondary school, potentially reducing effort required from parents later (Albornoz *et al.* (2018).

In this paper, I employ data from a detailed birth cohort study—the Millennium Cohort Study (MCS) (Platt *et al.* , 2014). The MCS follows a cohort of families from birth of the child, through ages 3, 5, 7, 11 and 14 years. This paper focuses on the primary school stage, until 11 years. The MCS collects detailed data on family background, their experience at school, skill development and parental involvement and investment. In terms of child outcomes, the MCS administers age-appropriate cognitive skill tests, and a well-validated scale measuring socio-emotional behaviours, allowing a characterisation of the development of both cognitive and non-cognitive skills over time. The MCS is eminently suitable for addressing the question of school allocations, because in addition to rich socio-economic and demographic data, the MCS also collects information on school choice and preferences for schools.

This study explores how missing out on a place at a preferred primary school affects pupils' cognitive and non-cognitive skill development, and traces the dynamics of parental behavioural responses. The challenge is to compare families who have a similar chance of admission to their preferred school, as well as having similar school preferences—yet one family misses out due to the nature of the admissions process.

In England, families submit a ranked list of schools they may wish their child to attend and places are allocated in accordance with these preferences by the local governing body who co-ordinate admissions. Oversubscribed schools are dealt with via a rule-based rationing mechanism: places are allocated based on observed characteristics of the child and their family. The priority criteria include whether they have special educational needs, whether the child lives in the school catchment area, and whether the child has an older sibling at the school. The tie-breaker above this is the straight-line distance of the child's home from the school. These criteria create some randomness in the chances of acceptance.

The MCS directly elicits the counterfactual school by asking parents, asking: if your current school was not your first choice school, which school was your first choice? From this question, we can compute the distance from a family's home address to their current school, and crucially, the distance to their *preferred school* (if they were not offered a place at their first-choice school).

Distance to preferred school, suitably interacted with geographical area, is especially useful for two reasons: (i) it informs on admission probability compared to other families in the local area and, (ii), distance to first-choice school also informs on as on preferences and unobserved family motivation. Previous research has shown that more ambitious families are willing to travel further in order to attend a high quality school, and this rate of substitution between travel-time and school quality varies based on ethnicity, economic status and the child's prior cognitive ability (Weldon, 2018).

Exploiting a estimation methods based on a conditional independence assumption (CIA), this paper compares the outcomes of children who have similar values of the observed variables—but did and did not miss out on a place a their preferred school. Documenting the *association* is a new contribution to this literature, and for a *causal* interpretation this strategy relies on a CIA assumption. Therefore I also assess the robustness of the results to deviations from the CIA via sensitivity analyses (Nannicini, 2007).

The findings do not reveal compelling evidence that missing out on a first choice school reduces cognitive development. There are small negative effects on Internalizing behaviours—an important component of the socio-emotional skill measure. However these fade out by age 11 years. In terms of parental responses, there are small (not statistically significant) increases in the probability of parental investment activities at ages 7—including helping with homework and paying for private tuition.

At age 11 years, those parents whose children missed out are *less* likely to help with homework than those who get into their preferred school, and significantly more likely now to pay for extra lessons and preparations for exams governing entry into selective schools. With the transition to secondary school imminent, using parental time to help with homework may be perceived by parents as an inferior substitute for private tutoring. These findings show that parental responses are important when considering policy impacts of education interventions.

3.2 Data and setting

3.2.1 Primary school admissions in England and Wales

For over 30 years (since the 1988 Education Reform Act), parents in England, Wales, and Northern Ireland[1] have been able to express a preference for the school they would like their child to attend. Parents may express, to their Local Authority (England and Wales) or Education Library Board (Northern Ireland), their choices of schools they wish for their child to attend (parents can select between three and six choice, depending on where they live). This paper focusses on England and Wales.

Places at oversubscribed schools are rationed based on observable characteristics: (1) children who are looked after by the state, (2) children with a Special Educational Needs certificate, (3) having older siblings at the school, and finally the distance between home

[1] In Scotland, families are allocated a place at their local school, but can then apply to the Local Education Authority to switch, if desired.

and school. In the case where two families live in precisely the same location, e.g., same apartment block, a coin flip is the decider. In the Millennium Cohort Study, about three quarters of primary schools were state controlled schools for which admissions are governed by the Local Authority in this way. The current system is rule-based, including proximity to school as a key determining factor. Compared to a completely random lottery, this method still encourages selection-by-mortgage, as wealthier families can move closer to their preferred school. This analysis takes the residential location decision as fixed.

3.2.2 Data

The data employed are from the Millennium Cohort Study (MCS), a nationally representative dataset of children and their families in the UK. The MCS target population comprises all children born between 1 September 2000 and 31 August 2001 (for England and Wales), and between 24 November 2000 and 11 January 2002 (for Scotland and Northern Ireland), alive and living in the UK at age 9 months, and eligible to receive child benefit at that age. The children are first surveyed aged 9 months, and followed up at age 3 years, 5 years, 7 years, 11 years and 14 years. The data collected encompass pregnancy, infant and early-childhood development, cognitive skills, behavioural problems, educational attainment, family life.

The study has a complex survey design, using clustering and stratification to ensure a sufficient sample size for analysis of target subgroups. The survey was clustered by electoral ward. Ethnic minority families, children living in deprived areas (bottom quartile of Child Poverty Index, a measure of the proportion of children in a ward whose families received means-tested benefits in the local area) and populations in Wales, Scotland and Northern Ireland were oversampled. The MCS attained a reasonable response rates; 72% in Wave 1. The patterns of participation are non-monotone: by the fifth survey, 54% of families had responded to all 5 waves, and another 20% had participated intermittently. Weights are provided in the MCS to account for attrition and the complex survey design, which I use throughout the analysis (specifically, the Wave 5 weights, because I use the sample which remains to Wave 5).

The survey principally tracks the child, and the bulk of the interview, especially in the early years is carried out with the main informant—typically the mother. A point of difference with the the MCS is that it also interviews fathers, about their life and involvement with the child and family. As the child grows older, they begin to conduct personal interviews directly with the child to elicit their experience in school and other development.

The majority of records are singleton births, yielding one cohort member per family, however out of the 18,552 children responding in Wave 1, there were also 208 twin births, and 11 triplets, in wave 1, yielding multiple cohort members per family in these cases. Following previous studies, I only use only the singleton births, to abstract from issues of birth order. Additionally, some of the key variables I use are only available for the first cohort member (e.g. distance to school). Given that the multiple birth sample is so small, it is highly unlikely that this exclusion would influence the results. Table 3.1 shows the structure of the MCS datasets.

3.2.3 Variables

Treatment variable

The MCS asked which school the cohort member currently attends, and whether this was their first choice school. The variable *Missed out* is equal to one if the current school is NOT the first school, and zero if it is the first choice school.

Cognitive development

The MCS administers a set of age appropriate tests of cognitive development: the British Ability Scales II (BAS II) (Elliott *et al.* , 1996), see Table 3.2. The BAS II has demonstrated construct validity as a measure of cognitive ability (Elliott *et al.* , 1997; Elliott, 1997) and high test-retest reliability (Elliott, 1997), making it suitable for a survey setting. At age 3, the BAS Naming Vocabulary and Bracken School Readiness tests are employed; at age 5, the BAS Verbal, Pattern Construction and Picture similarity tests are administered; at age 7 years, the Pattern Construction and Word Reading tests are used. Age 11 years used only the Verbal ability test is administered. The raw scores have been adjusted for the age-at-test and difficulty of the test.

The MCS cognitive ability scales are "tests of attainment based on the capability and motivation to complete a particular task under given conditions" (Platt *et al.* , 2014) —they not directly measure "general intelligence". In analyses of cognitive development, ability tests are typically combined using factor analysis, or related techniques, to approximate an underlying latent dimension of cognitive ability or general intelligence. Using multiple test measures helps reduce potential measurement error and regression to the mean, which could lead to spurious time trends (Jerrim & Vignoles, 2013). Following previous analyses of the MCS, e.g., Jones & Schoon (2008) and Bruckauf & Chzhen (2016), I use principal component analysis (PCA) to derive latent cognitive ability scores based on correlations between the observed age-adjusted test scores.

PCA is typically used in the social sciences as a data reduction device, reducing the variation in a set of correlated variables into a smaller set of components. The goal is to explain the largest amount of the variation in the data in the smallest number of components. The typical rule of thumb is to retain components if their associated eigenvalue is greater than 1, which was only the case for the first component. Therefore, I use the first component from the PCA, standardised to have mean zero and standard deviation of one.

Table 3.13 reports the estimates of factor loadings (the correlations between the latent factor and observed variables) at ages 3, 5 and 7 years. The latent factor accounted for 79 per cent of the underlying variance at age 3, 56 per cent at age 5 and 62 per cent at age 7. At age 11, there is only one cognitive ability test (BAS Verbal Similarity test). Therefore, only one measure of cognitive ability is used at age 11, and may be measured with more error than the combined measures used at earlier ages.

Socio-emotional development of the child

The Strengths and Difficulties Questionnaire (SDQ) is a well-validated screening tool for behavioural adjustment and mental health problems in children, and is widely used in both research and clinical practice. It been employed as a measure of the development of socio-emotional skills over time in the economics literature (Black & Kassenboehmer, 2017).

The SDQ is specifically designed for use among children, unlike the GHQ, which is primarily designed for use with adults. In the MCS, the SDQ is report by parents (about their child) when the children are aged 5, 7, 11 years, and by teachers when the children are aged 7 and 11 years. The SDQ comprises 25 items which generate an overall score, and can also been divided into five subscales, measuring different aspects of (dys)function. The subscales measure conduct problems, hyperactivity, emotional symptoms, peer problems and pro-social behavior.

Two commonly used summary measures of the SDQ are internalising and externalising scales. The internalising scale is the sum of emotional symptoms and peer problems; externalising problems is the sum of conduct problems and hyperactivity. These two summary measures have been found to have good properties in measuring childhood dysfunction. They are often added together to create a “Total difficulties score”.

Because the SDQ is conducted in all but the first of the MCS waves, when the children were very young, the development of any problems can be traced throughout childhood. In this paper I use both the the parent and teacher reports of the SDQ. As noted in (Black & Kassenboehmer, 2017), reports may in theory differ between parents and teachers—however, the teacher questionnaires are typically completed by a different teacher each wave, which is problematic for accurately measuring changes over time, as well as the teacher survey having a much survey response rate (57%).

Parental investments

In wave 3 and 4, the MCS asks whether anyone helps with the child with homework, across reading, writing and math, and the frequency (Not at all (=0), Less than a month, once or twice a month, once of twice a week, several times a week, every day) (=5). I created an additive variable which sums the frequency of help with homework across all subjects. In wave 5, the question is how often does anyone at home help with the cohort member’s homework (0=never or almost never, sometimes, usually, always=4), without reference to the specific subject.

The MCS also asks about whether the child has any extra private lessons to help with their schoolwork, creating the variable *Extra* indicating whether the parent has arranged for the child to receive extra lessons in either Math, Reading or Writing. This variable is available at ages 7 and 11 years. At age 11 years, the families are also asked specifically whether they have paid for private lessons to help them get into secondary school (*Extra prep*), and whether they are applying to a school with an entrance exam (*Entrance exam*).

3.2.4 Covariates

School selection criteria

The MCS asks parents first if they applied for school via the common LEA form (about 75%). For those who did apply via the LEA process, it asks which school their child attends, and whether the current school is the most preferred (about 94% report attending their most preferred school). If it is not the most preferred school, the survey asks which school was their first choice, and where applicable, which school was their second or third choice. From this information, a distance measure was constructed from home to the current school, for those for whom their current school was not preferred, we also have the distance to their first choice school (and second and third choice where appropriate).

The MCS data holds Ordnance Survey National Grid eastings and northings and uses these to map the MCS schools in a GIS. Because the Centre for Longitudinal Studies (CLS) also holds the unit postcodes of the addresses at interview of cohort members, it is possible to calculate the distances between these unit postcode centroids and the schools point data. These distances are straight line, or *as the crow flies*, based on straight line calculations. For this study, the distance measure has been banded into size bands to meet confidentiality requirements.

An indicator for whether the child has a certificate of Special Educational Needs (SEN) was taken from the age 7 parents report, which asks whether the child had a statement of special educational needs. There is also a teacher report at age 7 years, but the teacher survey has a lower response rate (57%) so this variable was not used in the main analyses. The Older Siblings module in the age 5 wave of the MCS asks about the older siblings of the main cohort member, including whether the older siblings attend the same school as the cohort member. In addition to variables which directly shape admissions probability, a range of detailed socio-economic and demographic variables are collected in the MCS:

- Gender of the child
- Ln(Household disposable income)
- Quintile of the Index of Multiple Deprivation
- Region (London, South West, South East, etc)
- Cognitive skills measured at age 3 years
- Month-of-birth of the child dummies
- Highest level of parental education
- Ethnicity (White / Indian / Pakistani / Bangladeshi / Black and Black British / Other or mixed)
- Parental religion

3.3 Empirical approach

3.3.1 Propensity Score Matching

The empirical approach used in this Chapter is situated in the potential outcomes framework. Let $Y_i(1)$ denote the value of the outcome Y under the treatment condition ($D = 1$) for individual i, and $Y_i(0)$ is the value of the outcome Y under the treatment condition ($D = 0$) for individual i. We wish to uncover the average difference in these outcomes among treated individuals: the Average Treatment Effect on the Treated (ATET) (Equation 3.1).

$$E[Y_i(1) - Y_i(0) \mid D_i = 1] \tag{3.1}$$

This analysis relies on a conditional independence assumption (Equation 3.2). This assumption states that the mean counterfactual outcome $Y(0)$ is independent of treatment status given a vector of covariates X_i.

$$E[Y_i(0) \perp D_i = 1, X_i] \tag{3.2}$$

This CIA assumptions means that the outcome under the control setting ($Y(0)$) is randomly assigned within groups defined by X_i, implied by Equation 3.3.

$$E[Y_i(0) \mid D_i = 1, X_i] = E[Y_i(0) \mid D_i = 0, X_i] \tag{3.3}$$

The credibility of this assumption depends heavily on the nature of X_i in relation to the research question at hand.[2] The second assumption, in addition to the CIA, is an overlap assumption (see Equation 3.3). Each treated unit should have a similar control unit for comparison.

$$Pr(D_i = 1 \mid X_i) < 1 \tag{3.4}$$

Given these assumptions, treatment assignment is effectively randomly assigned within strata defined by X_i. Therefore, the untreated outcomes can be used as counterfactual outcomes for the treated, who similar on the covariates X_i. A range of methods for measuring similarity between a treatment and control unit exist, including exact matching (matching units who have exactly the same values of X_i), coarsened exact matching (matching units within coarsened categories of X_i), multivariate distance matching (matching units with a similar value of a multivariate distance measure, e.g., Mahalanobis or Euclidean), propensity score matching (matching units with a similar conditional probability of being treated). Averaging these treatment-control comparisons, over the covariate distribution of the treatment group, gives an Average Treatment Effect on the Treated (ATET).

In most applications, X_i is high-dimensional, making it impractical to compare with exact strata defined by X_i or coarsened categories of X_i. A commonly used option which

[2]Note that no such restriction is placed on $Y_i(1)$ (outcome under treatment) or $Y_i(1) - Y_i(0)$ (selection-on-gains), as would required for and Average Treatment Effect.

resolves this issue is *propensity-score matching*—matching on the conditional probability of treatment $p(x) \equiv Pr(D = 1 \mid X)$, termed the *propensity score*. As shown in Rosenbaum (Rosenbaum & Rubin, 1983), if $Y(0)$ is independent of treatment given X_i, it will also be independent of treatment given $p(x)$. Hence, the propensity score can be used as a unidimensional summary of the observed conditioning variables.

Propensity score matching compares treated and control units with similar values of the estimated propensity score. How to select the control group matches give rise to various PSM estimation methods. For instance, nearest-neighbor or pairwise matching (selecting the closest n control units), radius matching (selecting all control units within a specified radius, or caliper), kernel matching (selecting all control units within a specified bandwidth, usually chosen via cross-validation, with the control units weighed in inverse proportion to their distance from the treated unit, governed by a kernel function).

Compared with Ordinary Least Squares regression, matching is a useful approach for at least three reasons. First, matching does not rely on a linear outcome model and it is less dependent on functional form. Often, many covariates are required to make the CIA assumption credible, and obtaining the right functional form may not be straightforward. Second, matching makes clear the comparison group, and facilitates an exploration of the common support and balance over the covariate distribution between treated and controls. That is, check if there is a control unit sufficiently similar to each treated unit. In contrast, "controlling" for covariates using OLS does not make clear the distribution of the control and treated units across covariates: in some cells there may not be a control unit to compare with a treated unit, and extrapolation or interpolation of the outcome is made based on the assumed functional form. It may also be hard to check whether the covariate distribution is similar between treatments and controls if there are many covariates, an issue which is resolved by using the propensity score.[3] Finally, matching yields an *population-averaged* treatment effect (e.g., an ATET or ATE), unlike OLS estimates which in general do not yield either an ATE or ATET (Angrist & Pischke, 2009; Gibbons *et al.* , 2014). OLS

[3]Inverse-probability-weighted regression adjustment (IPWRA) is another alternative to resolve this issue of covariate balance and functional form: by re-weighting the sample by the inverse probability of treatment, such that the distribution of X_i, via $p(x)$, is more similar across treatment and control groups, IPWRA method is then less reliant on the functional form imposed in the regression model.

is a minimum-variance estimator, and yields estimates which are implicitly weighted by the conditional variance of the treatment in each strata, penalising contributions from less precise strata and giving more weight to more precise strata.

Propensity score model and regression

In this application, the main estimation approach I employ is *kernel matching*. Kernel matching matches each treated unit to all control units within a specified distance (the "bandwidth", I use the default bandwidth of $p(x) = 0.06$ [4]), and gives a higher weight to control units closer to the treatment (governed by the kernel shape, I use a Epachnikov kernel). The standard errors are computed by bootstrapping.[5]

As a sensitivity check, I also employ multivariate distance matching on the Mahalanobis distance, fitting OLS models on the matched sample as a "bias adjustment". The rationale for this is as follows. Two advantages of matching are, first, *overlap* (ensuring the distribution of treatment and control units share a common support), and second, *balance*, (increasing the similarity of the distribution of covariates in the treatment and control groups).

Inexact matching, such as PSM, works to ensure balance. However, differences can remain, and overlaying regression on the matched sample acts as an insurance policy to soak up any remaining pre-treatment differences in covariates. In this context, combining matching and regression is referred to "bias correction". Another way to think about it to start from OLS: ensuring a more comparable sample by pre-processing the data to create a matched sample reduces the reliance on the correct functional form of the regression specification, especially in areas of the data with sparse common support.

A second reason to include the bias adjustment is to address the concerns highlighted in King & Nielsen (n.d.), who show that matching which randomly discards non-matches (e.g. nearest neighbor), in doing so reduces the sample size and can lead poor balance due to increased small-sample variation (among other arguments).

Covariate selection

[4] Paired with sensitivity analyses to deviations from this bandwidth)

[5] Unlike nearest neighbour matching, for which the analytic standard errors proposed by Alberto Abadie and colleagues are most appropriate (e.g., Abadie & Imbens (2008)), bootstrapped standard errors have not been proven to be invalid for kernel and local linear matching.

The selection of the covariates to include in the propensity score is motivated by the aim of making the CIA plausible. Whether the CIA is plausible depends on economic intuition and subject matter knowledge about the relationship between the X_i, and the potential values of Y_i. In the case of the allocation of school places, a number of factors shape both selection into treatment (missing out on a preferred school place) as well as outcomes. First, factors which determine admission to a school. Places are allocated based on a number of observable characteristics: (1) whether a child is looked after by the state, (2) whether the child has a certificate of special educational needs, (3) whether the child has an older sibling at the school, and (4) the straight line distance between home and school, which acts as a tie-breaker. As proxies for the admission criteria, I adjust for whether the parent reports that the child has a certificate of special educational needs, whether the child has any older siblings and distance to first-choice school. While there is a question on whether the child is in care, there are too few children in the sample reporting being in care to include this information.

Second, demographic and socio-economic characteristics also are associated with the likelihood of getting into a preferred school, the type of school which is preferred and child outcomes. Therefore, in addition to the admission criteria, I adjust for number of socio-economic variables, including log of household disposable income, region, ethnicity, quintiles of the Index of Multiple Deprivation, gender of the child, dummies for month-of-birth of the child, highest parental education. Comparing families in the same local area, and the same values of the selection criteria, is comparing families after residential sorting has already taken place; therefore this analysis does not pick up the total effect of response to the school choice mechanism. Residential choice is held fixed.

After selecting the covariates based on the economic considerations described above, I chose the model for the propensity score based on statistical goodness-of-fit criteria (i.e., what is the best fit of the probability of treatment), and ad-hoc consideration of the resulting balance statistics (as opposed to using an algorithm, such as in 'genetic' optimal matching algorithms, which include the balance statistics in the procedure for determining the propensity score specification). I include interactions between region and distance to preferred school, interaction between income and prior cognitive ability of the child and a quadratic in the Index of Multiple Deprivation.

3.3.2 Sensitivity analyses

To evaluate the sensitivity of the estimates to confounding, I employ the sensitivity analysis developed in Nannicini (2007). This sensitivity analysis simulates the effects of a potential binary confounder on the ATET. This method is similar in concept to many other sensitivity analyses in the statistics and econometrics literature, for example Altonji *et al.* (2008); Oster (2013) who also assess the sensitivity to unobserved confounding. The advantage of this specific method (Nannicini (2007)) is that is does not require a parametric outcome model for Y_i, making it suitable to use in a matching context.

The idea is that we suspect the conditional independence assumption (Equation 3.2) may not hold, given the X variables we observe. However, conditional on a omitted covariate U, the assumption would now plausibly hold (Equation 3.5).

$$E[Y_i(0) \perp D_i = 1, X_i, U_i] \tag{3.5}$$

Matching on U in addition to X allows us obtain a consistent estimate of the ATET.

By specifying the joint distribution of (U,D,Y) we can compute the "unbiased" ATT which accounts for the confounding effects of U, and compare this to our original "biased" estimate—which doesn't adjust for U—to assess the difference made by accounting for the unobserved covariate U.

To operationalise the method, one needs to specify the distribution of U in relation to D and Y.

$$p_{ij} \equiv Pr(U = 1 \mid D = i, Y = j) = Pr(U = 1 \mid D = i, Y = j, X) \tag{3.6}$$

Equation also highlights the maintained simplifying assumption that U is binary and independent of X. After specifying p_{ij}, the relevant value of U is assigned to each observation, depending on which category of i, j they are in, and U is included in the calculation of the ATET as an additional covariate. For a given set of parameters, the matching procedure is performed multiple times with varying draws of U, and the estimate of the ATET is the average over the estimate of the ATET in each simulation. The standard errors are calculated using Rubin's rules for computing standard errors across multiple datasets (i.e. the same procedure as used in multiple imputation for missing data).

I chose p_{ij} to mimic the distribution of ethnic minority status (equal to 0 for White families and 1 for non-White). Families of an ethnic minority are more likely to miss out on a preferred school, and have poorer child outcomes. The current literature shows that ethnic minority families are ambitious in their choice of school—they are willing to travel further to attend a high quality school. These beliefs and behavours are likely also correlated increased parental input and effort into their children's education—potentially acting as a confounder if this information is not fully captured in my observed data.

3.4 Findings

Table 3.3 shows little evidence for a difference in the probability of missing out on the first choice school by family income quintile or area deprivation quintile: the difference in the probability of getting into a first choice school is not significant across these categories. However, the right-hand panel of Table 3.3 does show differences in applying via the Local Authority process by income and area quintile. Affluent families in a higher income quintile, or quintile of area deprivation, are more likely to say they applied via LEA. However, this could be explained by differential recall bias or differential item-response. Table 3.4 shows the reasons stated by parents for their choice of school. Proximity and siblings are the most commonly cited reasons, partly perhaps due to the fact that families tend to base their location decision on the school desirability and the expected probability of getting in.

An initial question regarding the consequences of missing out on a first choice school, is whether the school quality is lower. While I do not have direct data on school characteristics, we can examine the satisfaction of parents across a range of domains of school performance. Table 3.5 reports parental satisfaction with their school by whether they got into their first choice or not. For example, 70% of parents who got their first choice are very satisfied, compared with 64% of those who missed out; a significant difference of 6 percentage points.

Table 3.7 reports summary statistics of the child outcomes by treatment status (missed out on first choice or not). There are significant differences in Internalising Problems (favouring those who get in), which appears to be driven by differences in its component subscale "Peer problems". This effect size fades out over time: no significant effects are detected at age 7 or 11 years. While those who miss out have slightly lower measured cognitive skills at each wave, these differences are not statistically significant. Table 3.8 reports summary statistics of the teacher report of the SDQ by treatment status (missed out on first choice or not). There are significant differences in Internalising Problems (favouring those who get in) in wave 4, and significant differences in Externalising Problems in wave 5. Table 3.9 summarises parental input by treatment status. Those who miss out are more likely to have help with homework and extra lessons, aside from at age 11 years when they seems to substitute from more homework to more paid lessons.

Table 3.10 shows "predetermined covariates" by treatment status; cohort members who have older siblings, and who live closer to their preferred school, are more likely to be offered a place, highlighting the importance of adjusting for these variables which shape admission probability.

Table 3.11 shows matching results comparing child outcomes at wave 3 by whether the child missed out on first choice school. In Wave 3, those who miss out have higher Internalising problems, but there is little difference in Externalising and cognitive skills. However, these differences are not statistically significant. In the subsequent Wave, there are no statistically significant differences in any of the skill measures considered.

Table 3.12 reports on matching results assessing the association of parental inputs with missing out on first choice school. There are positive associations between missing out and parental inputs in Wave 4, but they are not statistically significant. In Wave 5, the effects of missing out on private tuition and extra lessons for school entrance exams are positive and statistically significant: parents of children who miss out are more likely to invest in private tuition and exam preparations. Help with homework has a negative, but not statistically significant, coefficient. This may reflect substitution toward paid help, away from own inputs, now that the stakes are perceived to be high by parents.

The sensitivity analyses show that the positive association between missing out and increased parental inputs, at Wave 5, are unlikely to be fully explained by unobserved confounding. A potential confounder mimicking the distribution of minority ethnic status would reduce the estimated ATET for private tuition (extra lessons) from 0.084 to 0.074, a reduction of 13.5%. For preparatory lessons for selective school exams, the ATET reduces from 0.085 to 0.082, a reduction of 3.7%.

The Appendix Tables show further matching results, using nearest neigbour matching with a bias adjustment, rather than kernel matching, for comparison, which concord with the matching findings described above. No strong evidence is detected for effects on non-cognitive or cognitive development of missing out on a preferred primary school.

3.5 Concluding Remarks

School quality is an important factor shaping many life outcomes. This paper traces the skill development and parental responses of families who are not offered a place at their preferred primary school in England. Using a nationally representative cohort study, I compare the outcomes for families with similar preferences and probability of obtaining a place at their preferred school—however, some miss out on a place at their preferred school.

The results do not reveal evidence for detrimental effects of missing out on a first choice school on cognitive and non-cognitive skill development. In terms of parental responses, there are small, statistically insignificant increases in the probability of parental investment activities at ages 7—including helping with homework and paying for private tuition. At age 11 years, parents of children who miss out are significantly more likely to pay for extra lessons for their child. With the transition to secondary school imminent, using parental time to help with homework may be seen as an inferior substitute for private tutoring.

Further work remains to assess whether parental responses vary by demographic and socio-economic group. A policy concern is that while private tutoring is becoming a common tool to increase the chances of getting into selective schools, tutoring is more accessible to wealthy families—potentially increasing education inequalities. On the other hand, descriptive work shows that ethnic minority families, regardless of income, have a high propensity to take up private tutoring and other educational investments (Kirby, 2016). Differential parental responses could also have offsetting effects on the child's skill development, leading to the observed null average effect of missing out on cognitive and non-cognitive skills.

This study does not reveal any evidence that attending a non-preferred parental primary school has lasting effects on the skill development of children. However, the findings do highlight the importance of parental responses when considering policy impacts.

Table 3.1: Millennium Cohort Study data structure

Wave	Age of child	Number of children	Number of families
1	9 months	18,818	18,552
2	3 years	15,808	15,590
3	5 years	15,460	15,246
4	7 years	14,043	13,587
5	11 years	13,469	13,287

Table 3.2: Cognitive tests in each wave of Millennium cohort study

	Wave			
Type of test	MCS2	MCS3	MCS4	MCS5
BAS Naming Vocabulary	■	■		
Bracken School Readiness	■			
BAS Picture Similarity		■		
BAS Pattern Construction		■	■	
BAS Word Reading			■	
BAS Verbal Similarities				■
NFER Number Skills			■	
CANTAB Spatial Working Memory Task				■
CANTAB Cambridge Gambling Task				■

Table 3.3: Missed out by SES variables

Variable	Mean	SE	Mean	SE
	Missed out on first choice		Applied via LEA	
Income quintile				
Quintile 1 (low income)	0.06	0.00	0.66	0.01
Quintile 2	0.05	0.00	0.71	0.01
Quintile 3	0.05	0.00	0.78	0.01
Quintile 4	0.05	0.00	0.78	0.01
Quintile 5 (high income)	0.06	0.00	0.73	0.01
F-stat	2.84		29.63	
Area deprivation quintile				
Quintile 1 (most deprived)	0.06	0.00	0.67	0.00
Quintile 2	0.05	0.00	0.73	0.00
Quintile 3	0.05	0.00	0.74	0.00
Quintile 4	0.05	0.00	0.77	0.00
Quintile 5 (least deprived)	0.04	0.00	0.76	0.00
F-stat	2.32		19.83	

Notes: F-stat is the F-statistic for a joint test of equality of means across categories (e.g. testing the null hypothesis that the mean of Missed Out is equal across income quintiles

Table 3.4: Reasons for preferring chosen school

Main reason for choice	%	No.
Don't Know	0.0	1.0
School is near or nearest to home	58.8	4,603.0
His/her friends go or were intending to go there	3.3	260.0
His/her brother/sister went/go there	17.7	1,387.0
Other relative/parent went/go there	3.3	257.0
Wanted them to go to a different school to friend(s) or othe	0.1	9.0
Wanted them to go to a different school to brothers/sisters	0.1	8.0
Wanted them to go to a different school to other relative/pa	0.1	4.0
How likely it was that he/she would get a place	0.3	27.0
School has good exam results/academic reputation	7.1	555.0
General good impression of school	4.2	327.0
School has strong anti-bullying policy	0.1	4.0
School has small class sizes	0.7	54.0
School caters for special needs	0.4	29.0
School offers specialised curriculum e.g. music, dance, acti	0.1	6.0
School has good facilities	0.5	42.0
School offers childcare for parents who work or study	0.1	6.0
School is a feeder school	0.1	6.0
Religious grounds	1.2	95.0
Easy to get to on public transport	0.1	4.0
Ethnic mix of the school	0.1	10.0
School teaches in a language other than English	0.5	42.0
Wanted him/her to go to single-sex only school	0.0	1.0
Other reasons relating to the other children who go to the s	0.0	2.0
Other reason (specify)	1.1	90.0
Total	100.0	7,829.0

Table 3.5: Parental satisfaction with school in Wave 3

	Got in	Missed out	Difference	p-value
Parental satisfaction with school				
Very satisfied	0.70	0.64	0.06	0.02
Fairly satisfied	0.26	0.28	-0.02	0.34
Neither satisfied or dissatisfied	0.02	0.03	-0.01	0.05
Fairly dissatisfied	0.01	0.03	-0.01	0.04
Very dissatisfied	0.00	0.01	-0.01	0.12

Notes: Satisfaction of main parents with the child's school in Wave 3

Table 3.6: Child outcomes summary statistics by wave

	Mean	SD	Min	Max	N
Wave 3					
Internalising problems	2.50	2.54	0	18	9,152
Externalising problems	4.76	3.38	0	20	9,136
Conduct problems	1.49	1.49	0	10	9,196
Hyperactivity	3.27	2.36	0	10	9,141
Emotional problems	1.38	1.60	0	10	9,182
Peer problems	1.12	1.45	0	10	9,173
Cognitive ability	0.08	0.95	-4	3	9,290
Wave 4					
Internalising problems	2.79	2.79	0	19	8,045
Externalising problems	4.80	3.58	0	19	8,046
Conduct problems	1.41	1.54	0	10	8,075
Hyperactivity	3.39	2.51	0	10	8,051
Emotional problems	1.57	1.77	0	10	8,064
Peer problems	1.23	1.56	0	10	8,062
Cognitive ability	0.01	0.98	-3	2	8,124
Wave 5					
Internalising problems	3.30	3.23	0	19	7,583
Externalising problems	4.66	3.70	0	20	7,569
Conduct problems	1.47	1.67	0	10	7,592
Hyperactivity	3.19	2.48	0	10	7,573
Emotional problems	1.91	2.01	0	10	7,586
Peer problems	1.40	1.73	0	10	7,594
Cognitive ability	0.00	0.96	-4	2	7,681

Notes: Internalising problems is the sum of the emotional and peer problems subscales. Externalising problems is the sum of the conduct and hyperactivity subscales. Cognitive ability is the average over the British Ability Scale (BAS) measurements in that waves (BAS is an age appropriate measure of cognitive development.

Table 3.7: Child outcomes by treatment status

	Got in	Missed out	Difference	p-value	Raw N	Wtd N
Nobs	8,916.00	505.00				9,421
Wave 3						
Internalising problems	2.48	2.82	-0.34	0.01	9,152	10,161
Externalising problems	4.76	4.77	-0.01	0.96	9,136	10,153
Conduct problems	1.49	1.45	0.04	0.61	9,196	10,207
Hyperactivity	3.27	3.33	-0.05	0.63	9,141	10,157
Emotional problems	1.37	1.53	-0.15	0.05	9,182	10,191
Peer problems	1.11	1.30	-0.19	0.01	9,173	10,182
Cognitive ability	0.08	0.10	-0.02	0.58	9,290	10,246
Wave 4						
Internalising problems	2.78	2.88	-0.10	0.49	8,045	8,970
Externalising problems	4.79	4.85	-0.05	0.77	8,046	8,975
Conduct problems	1.41	1.45	-0.04	0.61	8,075	9,004
Hyperactivity	3.39	3.40	-0.01	0.96	8,051	8,979
Emotional problems	1.56	1.67	-0.11	0.26	8,064	8,992
Peer problems	1.23	1.22	0.00	0.95	8,062	8,987
Cognitive ability	0.01	0.03	-0.02	0.71	8,124	9,012
Wave 5						
Internalising problems	3.28	3.67	-0.39	0.04	7,583	8,415
Externalising problems	4.65	4.73	-0.08	0.70	7,569	8,408
Conduct problems	1.47	1.50	-0.03	0.73	7,592	8,421
Hyperactivity	3.19	3.24	-0.05	0.69	7,573	8,410
Emotional problems	1.90	2.08	-0.18	0.12	7,586	8,417
Peer problems	1.38	1.59	-0.20	0.03	7,594	8,424
Cognitive ability	-0.00	0.07	-0.07	0.22	7,681	8,475

Notes: Internalising problems is the sum of the emotional and peer problems subscales (as reported by the parent). Externalising problems is the sum of the conduct and hyperactivity subscales (as reported by the parent). Cognitive ability is the average over the British Ability Scale (BAS) measurements in that waves (BAS is an age appropriate measure of cognitive development. These figures are weighted using the survey weights, which account for the complex survey design and non-response. Standard errors are clustered by electoral ward.

Table 3.8: Child outcomes: as reported in the teacher survey by treatment status

	Got in	Missed out	Difference	p-value	Raw N	Wtd N
Wave 4						
Internalising problems (ts)	2.55	2.90	-0.35	0.06	5,355	5,773
Externalising problems (ts)	3.58	3.95	-0.37	0.12	5,357	5,775
Conduct problems	0.75	0.87	-0.12	0.21	5,357	5,775
Hyperactivity	2.83	3.08	-0.25	0.14	5,358	5,775
Emotional problems	1.41	1.67	-0.26	0.04	5,356	5,774
Peer problems	1.14	1.22	-0.09	0.38	5,355	5,773
Wave 5						
Internalising problems (ts)	2.59	2.71	-0.12	0.50	4,823.00	6,879
Externalising problems (ts)	2.96	3.59	-0.63	0.01	4,823.00	6,879
Conduct problems	0.69	0.91	-0.22	0.04	4,821.00	6,876
Hyperactivity	2.27	2.68	-0.41	0.01	4,820.00	6,876
Emotional problems	1.42	1.32	0.10	0.36	4,823.00	6,879
Peer problems	1.17	1.39	-0.22	0.05	4,821.00	6,872

Notes: Internalising problems is the sum of the emotional and peer problems subscales (as reported by the teacher). Externalising problems is the sum of the conduct and hyperactivity subscales (as reported by the teacher). These figures are weighted using the survey weights, which account for the complex survey design and non-response. Standard errors are clustered by electoral ward.

Table 3.9: Parental behaviours summary statistics

	Got in	Missed out	Difference	p-value	Raw N	Wtd N
Wave 3						
Help with homework	11.68	11.70	-0.02	0.89	9,421	10,370
Wave 4						
Help with homework	6.90	7.31	-0.41	0.07	8,278	8,777
Extra lessons	0.04	0.07	-0.03	0.02	8,237	8,731
Wave 5						
Help with homework	1.59	1.50	0.09	0.06	7,761	9,875
Extra lessons	0.21	0.34	-0.14	0.00	7,766	9,884
Did entrance exams?	0.16	0.33	-0.16	0.00	7,827	9,956
Lessons for entrance exams?	0.06	0.16	-0.09	0.00	7,828	9,957

Notes: These figures are weighted using the survey weights, which account for the complex survey design and non-response.

Table 3.10: Covariates from Wave 3

	Got in	Missed out	Difference	p-value	Raw N	Wtd N
Wave 3						
SEN	0.08	0.08	0.00	0.75	8,230	9,120
Older siblings?	0.84	0.72	0.12	0.00	9,420	10,369
Ethnicity						
White	0.90	0.81	0.09	0.00	9,420	10,369
Indian	0.02	0.03	-0.01	0.07	9,420	10,369
Pakistani	0.03	0.05	-0.02	0.07	9,420	10,369
Bangladeshi	0.01	0.02	-0.01	0.19	9,420	10,369
Black African or Caribbean	0.02	0.05	-0.03	0.00	9,420	10,369
Mixed, Other	0.02	0.04	-0.02	0.05	9,420	10,369
London	0.08	0.18	-0.10	0.00	9,419	10,366
Parental education						
None	0.03	0.03	-0.01	0.45	8,276	9,279
NVQ1	0.09	0.07	0.02	0.07	8,276	9,279
NVQ2	0.33	0.30	0.04	0.12	8,276	9,279
NVQ3	0.16	0.15	0.01	0.45	8,276	9,279
NVQ4	0.33	0.37	-0.03	0.18	8,276	9,279
NVQ5	0.05	0.09	-0.03	0.02	8,276	9,279
Distance from preferred school						
0-1 km	0.67	0.36	0.31	0.00	9,306	10,264
1-2 km	0.17	0.31	-0.14	0.00	9,306	10,264
2-5 km	0.11	0.20	-0.09	0.00	9,306	10,264
5-10 km	0.03	0.07	-0.03	0.00	9,306	10,264
10- 20 km	0.01	0.03	-0.02	0.01	9,306	10,264
20+ km	0.01	0.03	-0.02	0.00	9,306	10,264
Ln(Family income)	7.37	7.05	0.32	0.16	9,421	10,370

Notes: These figures are weighted using the survey weights, which account for the complex survey design and non-response.

Table 3.11: Effect of missing out on cognitive and non-cognitive skills from ages 5 to 11 years: PSM kernel matching (Epach kernal, bandwidth 0.06)

	ATT	s.e.	Total N
Wave 3			
Internalising problems	0.2477	0.1232	5,473
Externalising problems	0.0819	0.1831	5,473
Cognitive ability	-0.0154	0.0637	5,473
Wave 4			
Internalising problems	0.0271	0.1624	5,473
Externalising problems	0.1566	0.2343	5,473
Internalising problems (teacher survey)	0.2184	0.2389	5,473
Externalising problems (teacher survey)	0.2564	0.2884	5,473
Cognitive ability	0.0523	0.0610	5,473
Wave 5			
Internalising problems	0.1610	0.2337	5,473
Externalising problems	-0.0316	0.2233	5,473
Internalising problems (teacher survey)	0.1825	0.1845	5,473
Externalising problems (teacher survey)	0.0377	0.2410	5,473
Cognitive ability	0.0834	0.0532	5,473

Notes: Average Treatment Effect on the Treated (ATT) computed from a regression on a matched sample. Standard errors computed using 100 replications.

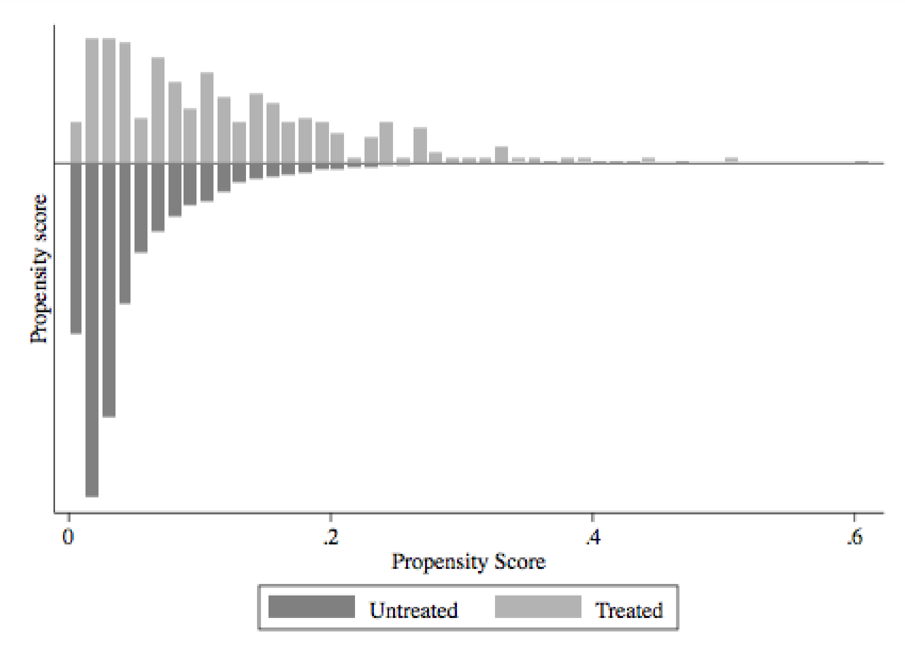
Table 3.12: Parental responses to missing out on cognitive and non-cognitive skills from ages 5 to 11 years: PSM kernel matching (Epachnikov kernel, bandwidth 0.06)

	ATT	s.e.	Total N
Wave 3			
Help with homework	0.0995	0.1770	5,473
Wave 4			
Extra lessons	0.0301	0.0210	5,473
Help with homework	0.1278	0.2648	5,473
Wave 5			
Extra lessons	0.0836	0.0311	5,473
Lessons for entrance exams?	0.0850	0.0182	5,473
Help with homework	-0.0189	0.0532	5,473

Notes: Average Treatment Effect on the Treated (ATT) computed from a regression on a matched sample. Standard errors computed using 100 replications.

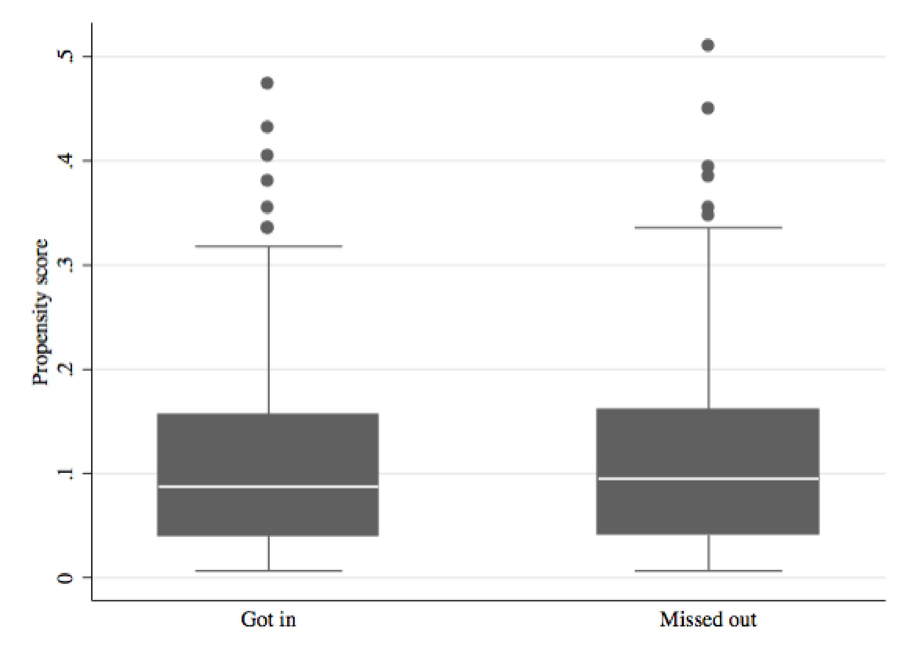
Appendix 3.A Matching validity tests

Figure 3.1: Histogram to check common support region



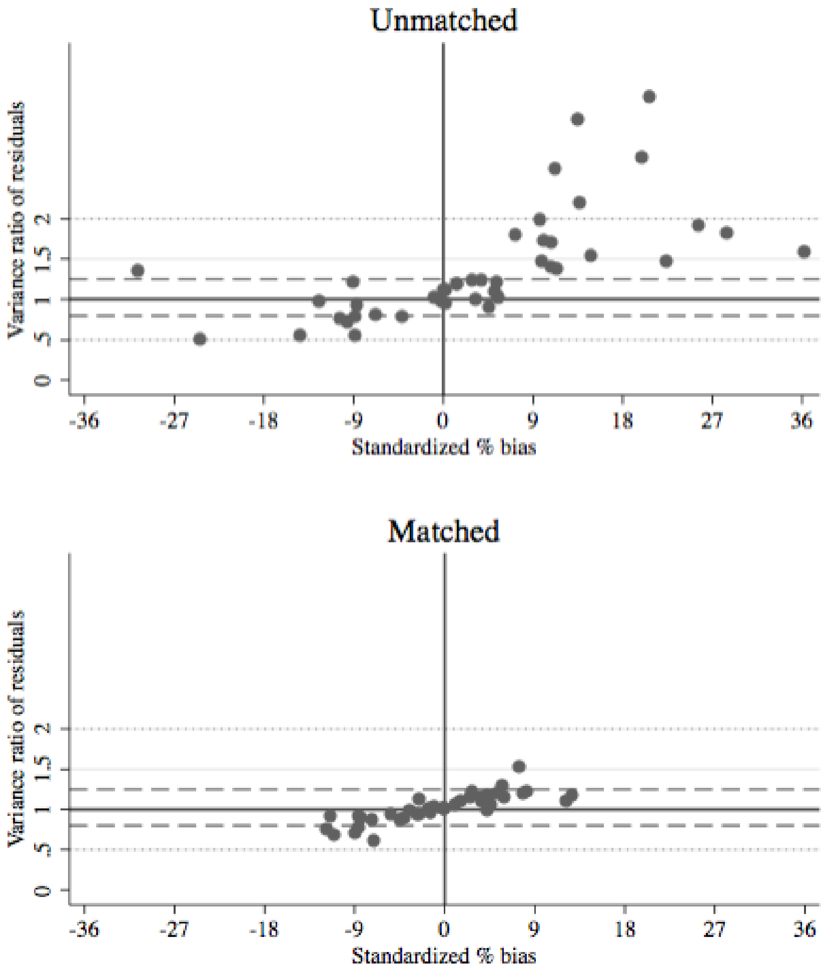
Notes: Histogram of the propensity score by treatment status.

Figure 3.2: Box plot to check common support region



Notes: Box plots of the propensity score by treatment status.

Figure 3.3: Scatter plot summarising balance statistics



Notes: Summary of balance statistics for propensity score matching generated by -psgraph-.

Appendix 3.B Principal components

Table 3.13: Factor loadings

Wave	Wave 2	Wave 3	Wave 4
Naming Vocabulary	0.71	0.57	-
Bracken School Readiness	0.71	-	-
Picture Similarity	-	0.59	-
pattern Construction	-	0.57	0.54
Word reading	-	-	0.56
Math	-	-	0.62
% explained	79.2	56.1	62.5

*Notes:*This tables reports on the factor loadings for each ability test contributing to the principal component analysis in each wave. The % explained report the % of variance explained by the first principal component in each wave.

Appendix 3.C Matching sensitivity

Table 3.14: Effect of missing out on cognitive and non-cognitive skills from ages 5 to 11 years: multivariate Mahalanobis distance kernel matching

	ATT	s.e.	Total N
Wave 3			
Internalising problems	0.3192	0.1710	5,341
Externalising problems	0.1420	0.2225	5,325
Cognitive ability	-0.0400	0.0514	5,297
Wave 4			
Internalising problems	0.0452	0.1955	5,296
Externalising problems	0.1309	0.2619	5,307
Internalising problems (teacher survey)	0.2161	0.2488	3,633
Externalising problems (teacher survey)	0.2489	0.3034	3,634
Cognitive ability	0.0448	0.0555	5,202
Wave 5			
Internalising problems	0.2074	0.2149	5,236
Externalising problems	-0.0206	0.2450	5,233
Internalising problems (teacher survey)	0.2005	0.2040	3,380
Externalising problems (teacher survey)	0.0902	0.3368	3,380
Cognitive ability	0.0684	0.0628	5,269

Notes: Average Treatment Effect on the Treated (ATT) computed from a regression on a matched sample. Standard errors computed using 100 replications.

Table 3.15: Parental responses to missing out on cognitive and non-cognitive skills from ages 5 to 11 years: multivariate Mahalanobis distance kernel matching

	ATT	s.e.	Total N
Wave 3			
Help with homework	0.2426	0.1883	5,341
Wave 4			
Extra lessons	0.0074	0.0168	5,341
Help with homework	-0.1188	0.3169	5,341
Wave 5			
Extra lessons	0.0507	0.0319	5,401
Lessons for entrance exams?	0.0515	0.0320	5,436
Help with homework	-0.0339	0.0288	5,400

Notes: Average Treatment Effect on the Treated (ATT) computed from a regression on a matched sample. Standard errors computed using 100 replications.

Conclusion

This thesis reports on new research assessing the role of schools in promoting the development of human capital. Using modern methods of policy evaluation, I explore the effects of two key aspects of education: (1) how many years a child spends in secondary school, and (2) attending a school preferred by parents.

In Chapter 1, I provide new evidence on the causal effects of basic education on later life outcomes—specifically, cognitive outcomes measured at approximately age 60 years. Basic education has been identified as a potential candidate in improving cognitive outcomes: for instance, based on observational evidence, *The Lancet*'s recent dementia Commission concluded that 8% of dementia cases could be avoided by increased levels of basic education (Livingston *et al.* , 2017). As successive governments have sought to address the problem of young people not in work, education or training (NEETs), the costs and benefits of different types of education at this crucial juncture in life is a fundamental input into policy analysis. Important parameters for policy design include the size of the causal effects of basic education, and the nature of the causal mechanisms through which such effects operate.

The findings in Chapter 1 suggest that an additional year of schooling have important *causal* effects on working memory, an important component of cognitive function. The analyses fail to detect evidence for causal effects on basic levels on numeracy, and verbal fluency. Staying on at school also reduces the probability of entering a manual or routine occupation, and results from a causal mediation analysis show that, in line with other studies, the occupation channel can explain up to about one-fifth of schooling's effects on memory. However, these figures are very imprecisely estimated.

While remaining agnostic about the role of occupation, I conclude that basic education causally improves an important component of cognitive function in older ages. These findings are consistent with the hypothesis that increased population levels of education could reduce the growth in burden of cognition-related disease, and support economic adjustment to a changing demographic structure.

Chapter 2 and 3 contribute to the area of school choice, tracing the effects of failing to gain a place at a preferred school. Chapter 2 assesses the long run effects of failing to gain a place at the secondary school most preferred by parents. The findings fail to detect consistent evidence that missing out on a place at a preferred school has deleterious effects on academic outcomes—holding socio-economic status, prior ability and residential local choice fixed. Schools are important in shaping more than just test scores, and the results reveal consistent evidence for increased engagement in risky behaviours and increasing rates of high school and university drop out, and reduced mental health among those who miss out on their preferred school.

The findings have a number of suggestive policy implications. Poor mental health casts a long shadow in life: mental health and well-being could prove a fruitful addition to the traditional measures of school quality in school league tables. Second, the nature of the assignment mechanism (how preferences are mapped to school allocations), and how parents understand and engage with this process, is important for ensuring equal access to schools. Finally, getting into a preferred school matters for important child outcomes. In this light, the oversubscription criteria which act as the gatekeeper to popular schools—determining who gets in and who misses out—is an opportune area for innovation to ensure fair access to schools.

Chapter 3 assesses the effects of missing out on a place at a preferred primary school, on cognitive and non-cognitive skill development and parental responses. The results provide scant evidence of a detrimental effect on skill development, but compared with those who get into their preferred school, parents do change their behaviours. Parents whose child misses out on a place are more likely increase their investments in the child through increasing private tutoring and selective school exam preparations, especially in the run-up to applying for secondary school.

This analyses adds important evidence to the nexus of the interaction between public and private investments in human capital. Behavioural responses are especially important in education settings, because the production of educational outcomes involves the interaction of pupils, parents, school teachers and administrator and the policy context. These types of interactions could dampen or accelerate policy interventions in unintended ways.

How to use scarce resources effectively to improve children and young people's life chances is an enduring research agenda. This thesis contributes new results to this important area.

Bibliography

Abadie, Alberto, & Imbens, Guido W. 2008. On the failure of the bootstrap for matching estimators. *Econometrica*, **76**(6), 1537–1557.

Abdulkadiroglu, A, & Sönmez, Tayfun. 2003. School Choice: A Mechanism Design Approach. *American Economic Review*, **93**(3), 729–747.

Abdulkadiroglu, Atila, Angrist, Joshua D., & Pathak, Parag A. 2014. The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica*, **82**(1), 137–196.

Abdulkadiroglu, Atila, Pathak, Parag, Schellenberg, Jonathan, & Walters, Christopher. 2017a. *Do Parents Value School Effectiveness?*

Abdulkadiroglu, Atila, Angrist, Joshua D., Narita, Yusuke, & Pathak, Parag A. 2017b. Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation. *Econometrica*, **85**(5), 1373–1432.

Abdulkadiroglu, Atila, Pathak, Parag A, & Walters, Christopher R. 2018. Free to Choose: Can School Choice Reduce Student Achievement? *American Economic Journal: Applied Economics*, **10**(1), 175–206.

Ahmadi-Abhari, Sara, Guzman-Castillo, Maria, Bandosz, Piotr, Shipley, Martin J., Muniz-Terrera, Graciela, Singh-Manoux, Archana, Kivimäki, Mika, Steptoe, Andrew, Capewell, Simon, O'Flaherty, Martin, & Brunner, Eric J. 2017. Temporal trend in dementia incidence since 2002 and projections for prevalence in England and Wales to 2040: modelling study. *BMJ*, **358**, 1–11.

Albornoz, Facundo, Berlinski, Samuel, & Cabrales, Antonio. 2018. Motivation, resources, and the organization of the school system. *Journal of the European Economic Association*, **16**(1), 199–231.

Almond, Douglas, & Mazumder, Bhashkar. 2013. Fetal Origins and Parental Responses. *Federal Reserve Bank of Chicago Economic Perspectives*.

Altonji, Joseph G, Elder, Todd E, & Taber, Christopher R. 2008. Using Selection on Observed Variables to Assess Bias from Unobservables when Evaluating Swan-Ganz Catheterization. *American Economic Review*, **98**(2), 345–350.

Ananth, Cande V, & VanderWeele, Tyler J. 2011. Placental abruption and perinatal mortality with preterm delivery as a mediator: disentangling direct and indirect effects. *American Journal of Epidemiology*, **174**(1), 99–108.

Andel, Ross, Silverstein, Merril, & Kareholt, Ingemar. 2015. The role of midlife occupational complexity and leisure activity in late-life cognition. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, **70**(2), 314–321.

Angrist, Joshua D., & Imbens, Guido W. 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, **90**(430), 431–442.

Angrist, Joshua David, & Pischke, Jorn-Steffen. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Autor, David, Figlio, David, Karbownik, Krzysztof, Roth, Jeffrey, & Wasserman, Melanie. 2016. *Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes*.

Banks, James, & Mazzonna, Fabrizio. 2012. The effect of education on old age cognitive abilities: evidence from a Regression Discontinuity Design. *The Economic Journal*, **122**(560), 418–448.

Banks, James, Carvalho, Leandro S., & Perez-Arce, Francisco. 2018. Education, Decision-making, and Economic Rationality. *Review of Economics and Statistics*.

Baron, Reuben M., & Kenny, David A. 1986. The moderator mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, **51**(6), 1173.

Bartalotti, Otávio C, & Brummet, Quentin O. 2016. Regression Discontinuity Designs with clustered data. *Advances in Econometrics*, **38**.

Beard, John R., Officer, Alana, & Cassels, Andrew. 2015. *World Report on Ageing and Health*. Tech. rept. World Health Organization, Geneva.

Becker, Gary S. 1964. Human capital theory. *Columbia, New York*.

Bernardi, Fabrizio, & Gratz, Michael. 2015. Making Up for an Unlucky Month of Birth in School: Causal Evidence on the Compensatory Advantage of Family Background in England. *Sociological Science*, **2**, 235–251.

Black, Nicole, & Kassenboehmer, Sonja C. 2017. Getting Weighed Down: The Effect of Childhood Obesity on the Development of Socioemotional Skills. *Journal of Human Capital*, **11**(2), 263–295.

Boyle, Patricia A., Yu, Lei, Wilson, Robert S., Gamble, Keith, Buchman, Aron S., & Bennett, David A. 2012. Poor decision making is a consequence of cognitive decline among older persons without alzheimer's disease or mild cognitive impairment. *PLoS ONE*, **7**(8), 5–9.

Brenøe, Anne Ardila, & Lundberg, Shelly. 2018. Gender gaps in the effects of childhood family environment: Do they persist into adulthood? *European Economic Review*, **109**(10313), 42–62.

Bruckauf, Zlata, & Chzhen, Yekaterina. 2016. Poverty and Children's Cognitive Trajectories: Evidence from the United Kingdom Millennium Cohort Study.

Brunello, Giorgio, Fort, Margherita, Schneeweis, Nicole, & Winter-Ebmer, Rudolf. 2015. The causal effect of education health: What is the role of health behaviours? *Health Economics*, jan.

Buckles, Kasey S., & Hungerman, Daniel M. 2013. Season of birth and later outcomes: Old questions, new answers. *Review of Economics and Statistics*, **95**(3), 711–724.

Burgess, Simon. 2015. Human Capital and Education : The State of the Art in the Economics of Education. 1–95.

Burgess, Simon, Greaves, Ellen, & Vignoles, Anna. 2017. *Understanding parental choices of secondary school in England using national administrative data*.

Calonico, Sebastian, Cattaneo, Matias D., & Farrell, Max H. 2016. rdrobust: Software for Regression Discontinuity Designs. 1–29.

Cantillon, Estelle. 2017. Broadening the market design approach to school choice. *Oxford Review of Economic Policy*, **33**(4), 613–634.

Case, Anne, & Deaton, Angus S. 2005. Broken down by work and sex: How our health declines. *Pages 185–212 of: Analyses in the Economics of Aging*. University of Chicago Press.

Case, Anne, & Paxson, Christina. 2009. Early life health and cognitive function in old age. *The American Economic Review*, **99**(2), 104–09.

Cattaneo, Matias D, Idrobo, Nicolas, & Titiunik, Rocío. 2017a. A Practical Introduction to Regression Discontinuity Designs. *Cambridge Elements: Quantitative and Computational Methods for Social Science-Cambridge University Press I*.

Cattaneo, Matias D, Titiunik, Rocío, & Vazquez-Bare, Gonzalo. 2017b. Comparing Inference Approaches in RD Designs: A Reexamination of the Effect of Head Start on Child Mortality. *Journal of Policy Analysis and Management*, **36**(3), 643–681.

Chêne, Geneviève, Dufouil, Carole, & Seshadri, Sudha. 2016. Incidence of dementia over three decades in the Framingham Heart Study. *New England Journal of Medicine*, **374**(6), 523–532.

Clark, Damon, & Royer, Heather. 2013. The effect of education on adult mortality and health: Evidence from Britain. *American Economic Review*, **103**(6), 2087–2120.

Coldron, John, Tanner, Emily, Finch, Steven, Shipton, Lucy, Wolstenholme, Claire, Willis, Ben, Demack, Sean, & Stiell, Bernadette. 2008. *Secondary School Admissions*.

Committee on Education and Skills. 2004. *Fourth Parliamentary Report on Education and Skills*.

Crawford, Claire, Dearden, Lorraine, & Greaves, Ellen. 2011. *Does when you are born matter? The impact of month of birth on children's cognitive and non-cognitive skills*.

Crespo, Laura, López-Noval, Borja, & Mira, Pedro. 2014. Compulsory schooling, education, depression and memory: New evidence from SHARELIFE. *Economics of Education Review*, **43**(dec), 36–46.

Crum, Rosa M, Anthony, James C, Bassett, Susan S, & Folstein, Marshal F. 1993. Population-based norms for the Mini-Mental State Examination by age and educational level. *Journal of the American Medical Association*, **269**(18), 2386–2391.

Cullen, Julie Berry, Jacob, Brian A., & Levitt, Steven. 2006. The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica*, **74**(5), 1191–1230.

de Walque, Damien. 2007. Does education affect smoking behaviors? *Journal of Health Economics*, **26**(5), 877–895.

Deary, Ian J., Corley, Janie, Gow, Alan J., Harris, Sarah E., Houlihan, Lorna M., Marioni, Riccardo E., Penke, Lars, Rafnsson, Snorri B., & Starr, John M. 2009. Age-associated cognitive decline. *British Medical Bulletin*, **92**(1), 135–152.

Deaton, Angus. 2010. Instruments, Randomization, and Learning about Development. *Journal of Economic Literature*, **48**(June), 424–455.

Deming, By David J, Hastings, Justine S, Kane, Thomas J, & Staiger, Douglas O. 2014. School Choice, School Quality and Post-Secondary Attainment. *American Economic Review*, **104**(3), 991–1013.

Department for Education. 2011. LSYPE User Guide to the Datasets : Wave 1 to Wave 7. 1–103.

Department for Education. 2017. *Unlocking Talent, fulfilling Potential*.

Department for Work and Pensions. 2013. *2010 to 2015 Government Policy: Older People*. Tech. rept.

Deuchert, Eva, Huber, Martin, & Schelker, Mark. 2018. Direct and indirect effects based on difference-in-differences with an application to political preferences following the Vietnam draft lottery. *Journal of Business & Economic Statistics*, 1–11.

DfE. 2001. *School Admissions Code 2001*.

Dickerson, Andy, Wilson, Rob, Kik, Genna, & Dhillon, Debra. 2012. *Developing Occupational Skills Profiles for the UK: A Feasibility Study*. Tech. rept. February. UK Comission for Employment and Skills.

Dickson, Matt, & Smith, Sarah. 2011. What determines the return to education: An extra year or a hurdle cleared? *Economics of Education Review*, **30**(6), 1167–1176.

Dickson, Matt, Gregg, Paul, & Robinson, Harriet. 2016. Early, late or never? When does parental education impact child outcomes? *The Economic Journal*, **126**(596), F184–F231.

Dippel, Christian, Gold, Robert, Heblich, Stephan, & Pinto, Rodrigo. 2017. *Instrumental Variables and Causal Mechanisms: Unpacking The Effect of Trade on Workers and Voters*.

Dobbie, Will, & Fryer, Roland 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), 58–75.

Dobbie, Will, & Fryer, Roland G. 2015. The Medium-Term Impacts of High-Achieving Charter Schools. *Journal of Political Economy*, **123**(5), 985–1037.

Dudovitz, RN, Chung, PJ, & Reber, S. 2018. Assessment of exposure to high-performing schools and risk of adolescent substance use: A natural experiment. *JAMA Pediatrics*, oct.

Elliott, C D, Smith, P, & McCulloch, K. 1996. British Ability Scales second edition (BAS II): administration and scoring manual. *London: NFER-Nelson*.

Elliott, COLIN D. 1997. The differential ability scales. *Contemporary intellectual assessment: Theories, tests, and issues*, 183–208.

Elliott, Colin D, Smith, Pauline, & McCulloch, Kay. 1997. British ability scales second edition (BAS II): Technical manual. *Windsor: NFER-Nelson*.

Elsner, Benjamin, & Isphording, Ingo E. 2017. A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics*, **35**(3), 787–828.

Fisher, Gwenith G., Stachowski, Alicia, Infurna, Frank J., Faul, Jessica D., Grosch, James, & Tetrick, Lois E. 2014. Mental work demands, retirement, and longitudinal trajectories of cognitive functioning. *Journal of Occupational Health Psychology*, **19**(2), 231–242.

Flatley, John, Connolly, Helen, Higgins, Vanessa, Williams, John, Coldron, John, Stephenson, Kathy, & Logie, Angela. 2001. Parents ' experiences of the process of choosing a secondary school. *Department for Education and Skills*, 3–267.

Flynn, James R. 1987. Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, **101**(2), 171–191.

Fortin, Nicole, Lemieux, Thomas, & Firpo, Sergio. 2011. Decomposition methods in economics. *Pages 1–102 of: Handbook of labor economics*, vol. 4. Elsevier.

Fredriksson, Peter, Öckert, Björn, & Oosterbeek, Hessel. 2016. Parental responses to public investments in children: Evidence from a maximum class size rule. *Journal of Human Resources*, **51**(4), 832–868.

Fujishiro, Kaori, MacDonald, Leslie A, Crowe, Michael, McClure, Leslie A, Howard, Virginia J, & Wadley, Virginia G. 2017. The Role of Occupation in Explaining Cognitive Functioning in Later Life: Education and Occupational Complexity in a U.S. National Sample of Black and White Men and Women. *The Journals of Gerontology: Series B*, **00**(00), 1–11.

Gelman, Andrew, & Zelizer, Adam. 2015. Evidence on the deleterious impact of sustained use of polynomial regression on causal inference. *Research & Politics*, **2**(1), 2053168015569830.

Gibbons, Charles E, Serrato, Juan Carlos Suárez, & Urbancic, Michael B. 2014. Broken or fixed effects? *Journal of Econometric Methods*.

Gibbons, Stephen, & Machin, Stephen. 2003. Valuing English primary schools. *Journal of Urban Economics*, **53**(2).

Gibbons, Stephen, & Silva, Olmo. 2011. School quality, child wellbeing and parents' satisfaction. *Economics of Education Review*, **30**(2), 312–331.

Glymour, M. Maria, Kawachi, Ichiro, Jencks, Christopher S., & Berkman, Lisa F. 2008. Does childhood schooling affect old age memory or mental status? Using state schooling laws as natural experiments. *Journal of Epidemiology and Community Health*, **62**(6), 532–537.

Goldberger, Arthur S. 1972. Structural equation methods in the social sciences. *Econometrica: Journal of the Econometric Society*, 979–1001.

Grimard, Franque, & Parent, Daniel. 2007. Education and smoking: were Vietnam war draft avoiders also more likely to avoid smoking? *Journal of Health Economics*, **26**(5), 896–926.

Grossman, Michael. 1972. On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, **80**(2), 232–255.

Haavelmo, Trygve. 1943. The statistical implications of a system of simultaneous equations. *Econometrica, Journal of the Econometric Society*, 1–12.

Hahn, Jinyong, Todd, Petra E., & van der Klaauw, Wilbert. 2001. Identification and estimation of treatment effects with a Regression-Discontinuity Design. *Econometrica*, **1**(69), 201–209.

Hansen, Kirstine. 2014. Moving house for education in the pre-school years. *British Educational Research Journal*, **40**(3), 483–500.

Harmon, Colm, & Walker, Ian. 1995. Estimates of the economic return to schooling for the United Kingdom. *The American Economic Review*, **85**(5), 1278–1286.

Heckman, James J. 1995. Lessons from the Bell Curve. *Journal of Political Economy*, **103**(5), 1091–1120.

Heckman, James J. 2006. The technology and neuroscience of capacity formation. *PNAS*, **I**(1), 1–8.

Heckman, James J, & Pinto, Rodrigo. 2015. Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric reviews*, **34**(1-2), 6–31.

Heckman, James J., Stixrud, Jora, & Urzua, Sergio. 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*.

Henderson, Morag. 2017. Next Steps User Guide (First Edition) Edited by Lisa Calderwood.

Herrnstein, Richard J., & Murray, Charles. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Hoekstra, Mark, Mouganie, Pierre, & Wang, Yaojing. 2018. Peer Quality and the Academic Benefits to Attending Better Schools. *Journal of Labor Economics*, **36**(4), 881–884.

Hsu, Joanna, & Willis, Robert. 2013. Dementia Risk and Financial Decision Making by Older Households: The Impact of Information. *Journal of Human Capital*, **45**, 1–40.

Huang, Wei, Lei, Xiaoyan, Ridder, Geert, Strauss, John, & Zhao, Yaohui. 2013. Health, Height, Height Shrinkage, and SES at Older Ages: Evidence from China. *American Economic Journal: Applied Economics*, **5**(2), 86–121.

Huber, Martin. 2015. Causal Pitfalls in the Decomposition of Wage Gaps. *Journal of Business & Economic Statistics*, **33**(2), 179–191.

Huber, Martin. 2016. Disentangling policy effects into causal channels. *IZA World of Labor*.

Huber, Martin, Lechner, Michael, & Steinmayr, Andreas. 2015. Radius matching on the propensity score with bias adjustment: tuning parameters and finite sample behaviour. *Empirical Economics*, **49**(1), 1–31.

Huber, Martin, Lechner, Michael, & Mellace, Giovanni. 2016. The finite sample performance of estimators for mediation analysis under sequential conditional independence. *Journal of Business & Economic Statistics*, **31**(1), 139–160.

Huber, Martin, Frölich, Markus, & Huber, Martin. 2017a. Direct and indirect treatment effects: Causal chains and mediation analysis with instrumental variables. *Journal of the Royal Statistical Society Series B.*, **79**(5), 645–1666.

Huber, Martin, Lechner, Michael, & Mellace, Giovanni. 2017b. Why Do Tougher Caseworkers Increase Employment? The Role of Program Assignment as a Causal Mechanism. *Review of Economics and Statistics*, **99**(1), 180–183.

Imai, Kosuke, Keele, Luke, Tingley, Dustin, & Yamamoto, Teppei. 2011. Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review*, **105**(04), 765–789.

Imbens, Guido W., & Lemieux, Thomas. 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, **142**(2), 615–635.

Institute for Social and Economic Research, & NatCen Social Research. 2015. *Understanding Society: Waves 1-5, 2009-2014: Special Licence Access. 6th Edition. SN: 6931*.

Jackson, C Kirabo, & Beuermann, Diether W. 2018. *Do Parents Know Best? The Short and Long-Run Effects of Attending the Schools that Parents Prefer*.

Jann, Ben. 2017. *KMATCH: Stata module for multivariate-distance and propensity-score matching*.

Jerrim, John, & Vignoles, Anna. 2013. Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **176**(4), 887–906.

Jones, Andrew M., Rice, Nigel, & Rosa Dias, Pedro. 2011. Quality of schooling and inequality of opportunity in health. *Empirical Economics*, **42**(2), 369–394.

Jones, Elizabeth M, & Schoon, Ingrid. 2008. Child cognition and behaviour. *Millennium Cohort Study Third Survey: A user's guide to initial findings*, 118–144.

Keele, Luke, Tingley, Dustin, & Yamamoto, Teppei. 2015. Identifying mechanisms behind policy interventions via causal mediation analysis. *Journal of Policy Analysis and Management*, **34**(4), 937–963.

Kim, Jee Wook, Lee, Dong Young, Seo, Eun Hyun, Sohn, Bo Kyung, Choe, Young Min, Kim, Shin Gyeom, Park, Shin Young, Choo, I L Han, Youn, Jong Chul, Jhoo, Jin Hyeong, Kim, Ki Woong, & Woo, Jong Inn. 2014. Improvement of screening accuracy of Mini-Mental State Examination for mild cognitive impairment and Non-Alzheimer's Disease Dementia by supplementation of verbal fluency performance. *Psychiatry Investigation*, **11**(1), 44–51.

King, Gary, & Nielsen, Richard. *Why Propensity Scores Should Not Be Used for Matching*.

Kirby, Philip. 2016. Shadow Schooling: private tuition and social mobility in the UK.

Lee, David S., & Card, David. 2008. Regression discontinuity inference with specification error. *Journal of Econometrics*, **142**(2), 655–674.

Livingston, Gill, Sommerlad, Andrew, Orgeta, Vasiliki, Costafreda, Sergi G., Huntley, Jonathan, Ames, David, Ballard, Clive, Banerjee, Sube, Burns, Alistair, Cohen-Mansfield, Jiska, Cooper, Claudia, Fox, Nick, Gitlin, Laura N., Howard, Robert, Kales, Helen C., Larson, Eric B., Ritchie, Karen, Rockwood, Kenneth, Sampson, Elizabeth L., Samus, Quincy, Schneider, Lon S., Selbæk, Geir, Teri, Linda, & Mukadam, Naaheed. 2017. Dementia prevention, intervention, and care. *The Lancet*.

Lochner, Lance. 2011. *Non-production benefits of education: Crime, health and good citizenship*.

Mazumder, Bhashkar. 2012. The effects of education on health and mortality. *Nordic Economic Policy Review*.

Mazzonna, Fabrizio. 2014. The long lasting effects of education on old age health: evidence of gender differences. *Social Science & Medicine*, **101**(jan), 129–38.

Mazzonna, Fabrizio, & Peracchi, Franco. 2012. Ageing, cognitive abilities and retirement. *European Economic Review*, **56**(4), 691–710.

Mazzonna, Fabrizio, & Peracchi, Franco. 2014. Unhealthy Retirement? *Journal of Human Resources*, **52**(1), 128–151.

McFall, Stephanie. 2013a. Understanding Society the UK Household longitudinal study: Cognitive ability Measures. *Understanding Society*.

McFall, Stephanie. 2013b. Understanding Society The UK Household Longitudinal Study Waves 1-3, User Manual. *Colchester: University of Essex*.

Mendolia, Silvia, Paloyo, Alfredo R, Walker, Ian, & Mendolia, Silvia. 2018. Heterogeneous Effects of High School Peers on Educational Outcomes. *Oxford Economic Papers*, **70**(3), 613–634.

Muurinen, Jaana-Marja. 1982. Demand for health: A Generalised Grossman model. *Journal of Health Economics*, **1**, 5–28.

Muurinen, Jaana-Marja, & Grand, Julian. 1985. The economic analysis of inequalities in health. *Social science & medicine*, **20**(10), 1029–1035.

Nannicini, Tommaso. 2007. Simulation-based sensitivity analysis for matching estimators. *Stata Journal*, **7**(3).

OECD. 2006. *Live Longer, Work Longer*.

Oreopoulos, Philip. 2008. Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, **96**(1), 152–175.

Oreopoulos, Philip, & Salvanes, Kjell G. 2011. Priceless: The nonpecuniary benefits of schooling. *Journal of Economic Perspectives*, **25**(1), 159–184.

Oster, Emily. 2013. *Unobservable selection and coefficient stability: Theory and evidence*.

Park, Soojin, & Kürüm, Esra. 2018. Causal mediation analysis with multiple mediators in the presence of treatment noncompliance. *Statistics in Medicine*, **37**(11), 1810–1829.

Pathak, Parag A., & Sönmez, Tayfun. 2013. School admissions reform in Chicago and England: Comparing mechanisms by their vulnerability to manipulation. *American Economic Review*, **103**(1), 80–106.

Plassman, Brenda L., Welsh, Kathleen. A., Helms, M., Brandt, Jason, Page, William F., & Breitner, John C. S. 1995. Intelligence and education as predictors of cognitive state in late life: A 50-year follow-up. *Neurology*, **45**(8), 1446–1450.

Platt, Lucinda, Smith, Kate, Parsons, Samantha, Connelly, Roxanne, Joshi, Heather, Rosenberg, Rachel, Hansen, Kirstine, Brown, Matt, Sullivan, Alice, Chatzltheocharl, Stella, & Others. 2014. Millennium Cohort Study: initial findings from the Age 11 survey.

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

Prince, Martin, Knapp, Martin, Guerchet, Maëlenn, & McCrone, Paul. 2014. *Dementia UK: Update*. Tech. rept. 1. Alzheimer's Society, London.

Restrepo, Brandon J. 2016. Parental investment responses to a low birth weight outcome: who compensates and who reinforces? *Journal of Population Economics*, **29**(4), 969–989.

Rohwedder, Susann, & Willis, Robert J. 2010. Mental retirement. *Journal of Economic Perspectives*, **24**(October 2009), 1–19.

Rosenbaum, Paul R, & Rubin, Donald B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**(1), 41–55.

Roth, M., Tym, E., Mountjoy, C. Q., Huppert, Felicia A., Hendrie, H., Verma, S., & Goddard, R. 1986. CAMDEX. A standardised instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. *The British Journal of Psychiatry*, **149**(6), 698–709.

Schneeweis, Nicole, Skirbekk, Vegard, & Winter-Ebmer, Rudolf. 2014. Does education improve cognitive performance four decades after school completion? *Demography*, **51**(2), 619–643.

Skirbekk, Vegard, Stonawski, Marcin, Bonsang, Eric, & Staudinger, Ursula M. 2013. The Flynn effect and population aging. *Intelligence*, **41**(3), 169–177.

Solon, Gary, Haider, Steven J., & Wooldridge, Jeffrey M. 2015. What Are We Weighting For? *Journal of Human Resources*, **50**(2), 301–316.

Stern, Yaakov. 2002. What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, **8**(03), 448–460.

Tchetgen, Eric J Tchetgen, Lin, Sheng Hsuan, Tchetgen Tchetgen, Eric J, Lin, Sheng Hsuan, Tchetgen, Eric J Tchetgen, & Lin, Sheng Hsuan. 2012. *Robust Estimation of Pure / Natural Direct Effects with Mediator Measurement Error*.

Tiebout, Charles M. 1956. A pure theory of local expenditures. *Journal of Political Economy*, **64**(5), 416–424.

Tucker-Drob, Elliot M. 2011. Neurocognitive Functions and Everyday Functions Change Together in Old Age. *Neuropsychology*, **25**(3), 368–377.

University College London, UCL Institute of Education Centre for Longitudinal Studies. 2018. *Next Steps: Sweeps 1-8, 2004-2016: Secure Access. [data collection]. 4th Edition. UK Data Service. SN: 7104*.

VanderWeele, T J. 2015. *Explanation in causal inference: Methods for mediation and interaction*. Oxford University Press.

VanderWeele, Tyler J. 2010. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology (Cambridge, Mass.)*, **21**(4), 540–551.

Weldon, Matthew. 2018. *Secondary school choice and selection: insights from new nationa preferences data*. Tech. rept. Lancaster University.

West, Anne, Hind, Audrey, & Pennell, Hazel. 2004. School admissions and 'selection' in comprehensive schools: Policy and practice. *Oxford Review of Education*, **30**(3), 347–369.

West, Anne, Barham, Eleanor, & Hind, Audrey. 2011. Secondary school admissions in England 2001 to 2008: Changing legislation, policy and practice. *Oxford Review of Education*, **37**(1), 1–20.

Whalley, Lawrence J., Deary, Ian J., Appleton, Charlotte L., & Starr, John M. 2004. Cognitive reserve and the neurobiology of cognitive aging. *Ageing Research Reviews*, **3**(4), 369–382.

Williams, John, Coldron, John, Fearon, Jane, Stephenson, Kathy, Logie, Angela, & Smith, Nicola. 2001. An analysis of the policies and practices of admission authorities in England. *BERA Conference proceedings*.

Woodin, Tom, McCulloch, Gary, & Cowan, Steven. 2012. Raising the participation age in historical perspective: policy learning from the past? *British Educational Research Journal*, 1–19.

Wraw, Christina, Deary, Ian J., Gale, Catharine R., & Der, Geoff. 2015. Intelligence in youth and health at age 50. *Intelligence*, **53**(nov), 23–32.

Wu, Qiong, Tchetgen, Eric J Tchetgen, Osypuk, Theresa L., White, Kellee, Mujahid, Mahasin, & Glymour, M. Maria. 2013. Combining direct and proxy assessments to reduce attrition bias in a longitudinal study. *Alzheimer Disease and Associated Disorders*, **27**(3), 207.

Yamamoto, Teppei. 2014. *Identification and estimation of causal mediation effects with treatment noncompliance*.